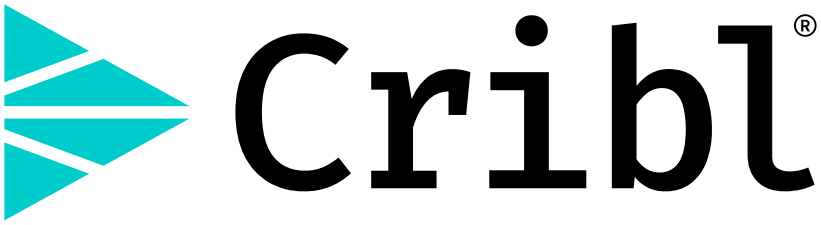


Cribl Edge Documentation Manual

Version: v4.6

Generated: 2024-04-17 06:58:20

1. Introduction

1.1. About Cribl Edge

All the docs to goat you started with Cribl Edge.

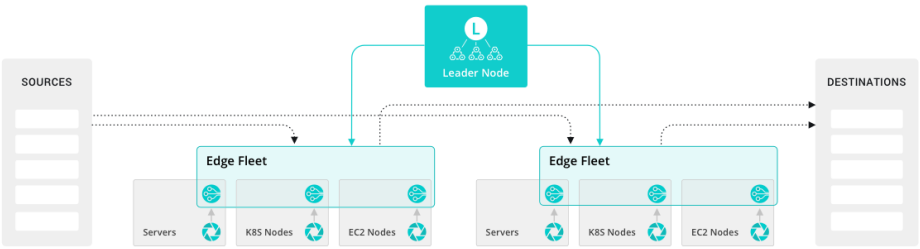
Download all docs as a PDF – v.4.5.1 | Download Binary | Edge on Cribl.Cloud

- Questions not answered here? We'd love to help you. Meet us in #Cribl Community Slack – sign up here.

What Is Cribl Edge?

Cribl Edge helps you collect and process observability data – logs, metrics, application data, etc. – in real time, from your Linux and Windows machines, apps, microservices etc., and deliver them to Cribl Stream or any supported destination.

- Automatic discovery of host, container and application metrics, logs, etc. on endpoints.
- Great Getting Data In (GDI) experience: explore, preview, build configs etc.
- Collect, Process and Forward
    - Collect logs, metrics (host, process, container), traces (coming soon), custom command output etc.
    - All the processing power of functions & pipelines at your fingertips
    - Forward data to any of the many supported destinations
- Centralized Configuration & Management
    - Visual configuration authoring, version controlled etc.
    - Teleportation-to-the-edge for local preview and configuration validation



Edge Leader, Fleets, and Destinations

Who Is Cribl Edge For?

Cribl Edge is built for ITOps and SecOps Engineers and users of operational and security intelligence products and services who want to simplify the process of collecting and analyzing data from a host.

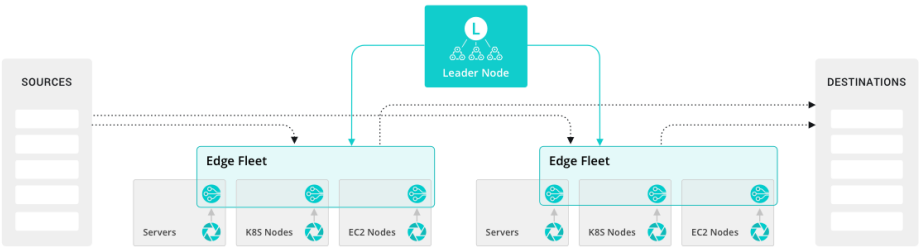
Cribl University offers two courses that provide a good overview of Cribl Edge: The What & Why of Edge of Cribl Edge and Key Concepts.To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

1.2. Basic Concepts

Notable features and concepts to get a fundamental understanding of Cribl Edge

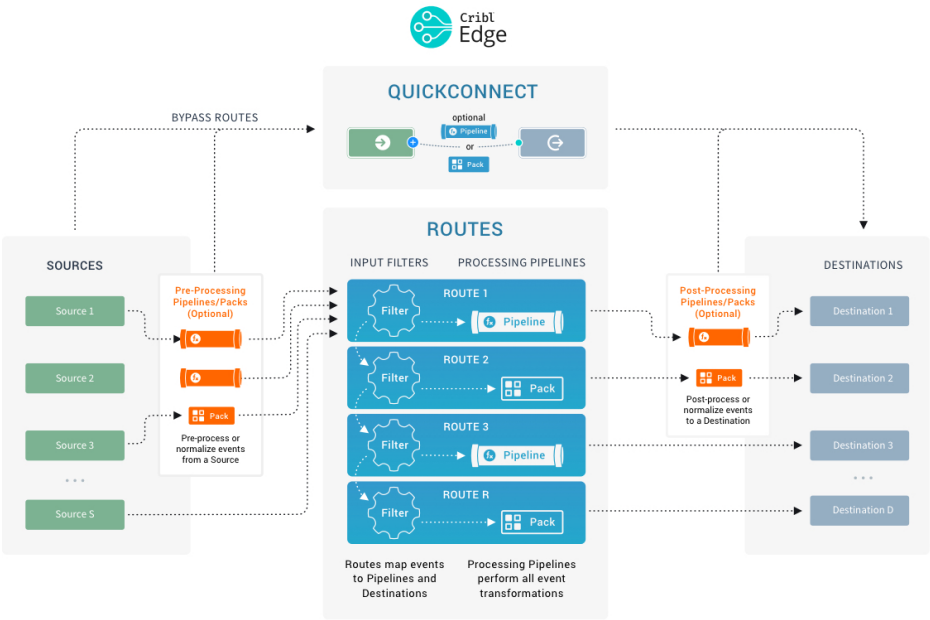
---

As we describe features and concepts, it helps to have a mental model of Cribl Edge as a system that collects and process observability data – logs, metrics, application data, etc. – in near real time, from your Linux and Windows machines, apps, microservices etc., and delivers them to Cribl Stream or any supported destination.



Edge processing, management, and receivers

Let's zoom in on the center of the above diagram, into an Edge Node, to take a closer look at the processing and transformation options. The basic interface concepts to work with are Sources, which collect data; and Routes, which manage data flowing through Pipelines, which consist of Functions.



Routes, Pipelines, Functions

Sources

Sources are configurations that enable Edge to collect data from the local machine (logs, metrics, etc.) or to receive data from remote senders (TCP, Syslog etc.).

QuickConnect

QuickConnect is a graphical interface for setting up data flow through your Edge deployment. You can quickly drag and drop connections between Sources and Destinations, optionally including – or excluding – Pipelines or Packs.

The only major constraint is that QuickConnect completely bypasses Routes. So QuickConnect configurations have no Routing table, and no conditional cloning, cascading, or routing of data – every QuickConnect connection is parallel and independent.

You can access the **QuickConnect** interface from the top nav's **Collect** screen.

Routes

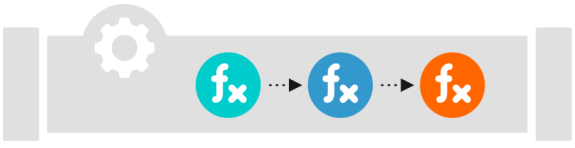
Routes evaluate incoming events against **filter** expressions to find the appropriate Pipeline to send them to. Routes are **evaluated in order**. Each Route can be associated **with only one** Pipeline and one output (configured as a Edge **Destination**).

By default, each Route is created with its `Final` flag set to `Yes`. With this setting, a Route-Pipeline-Destination set will consume events that match its filter, and that's that.

However, if you disable the Route's `Final` flag, one or more event **clones** will be sent down the associated Pipeline, while the original event continues down Edge's Routing table to be evaluated against other configured Routes. This is very useful in cases where the same set of events needs to be processed in multiple ways, and delivered to different destinations. For more details, see Routes.

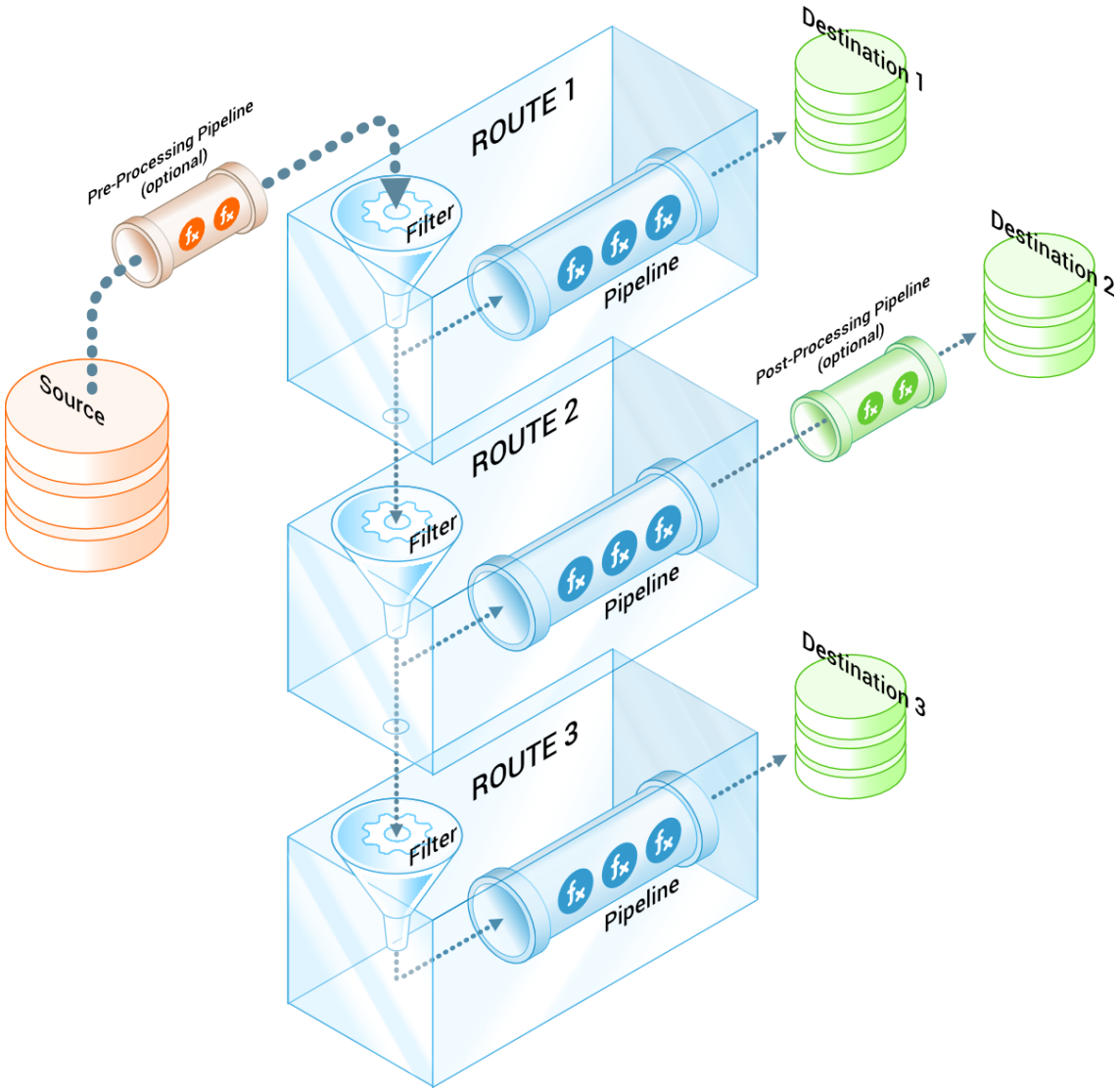
Pipelines

A series of Functions is called a Pipeline, and the order in which you specify the Functions determines their execution order. Routes deliver Events to the beginning of a Pipeline, and as they're processed by a Function, the events are passed to the next Function down the line.



Pipelines attached to Routes are called **processing Pipelines**. You can optionally attach **pre-processing Pipelines** to individual Edge Sources, and attach **post-processing Pipelines** to Edge Destinations. All

Pipelines are configured through the same UI – these three designations are determined only by a Pipeline's placement in Edge's data flow.



Pipelines categorized by position

Events only move forward – toward the end of a Pipeline, and eventually out of the system. For more details, see Pipelines.

Functions

At its core, a **Function** is a piece of code that executes on an event, and that encapsulates the smallest amount of processing that can happen to that event. For instance, a very simple Function can be one that replaces the term `foo` with `bar` on each event. Another one can hash or encrypt `bar`. Yet another function can add a field – say, `dc=jfk-42` – to any event with `source=*us-nyc-application.log`.

Functions stacked in a Pipeline

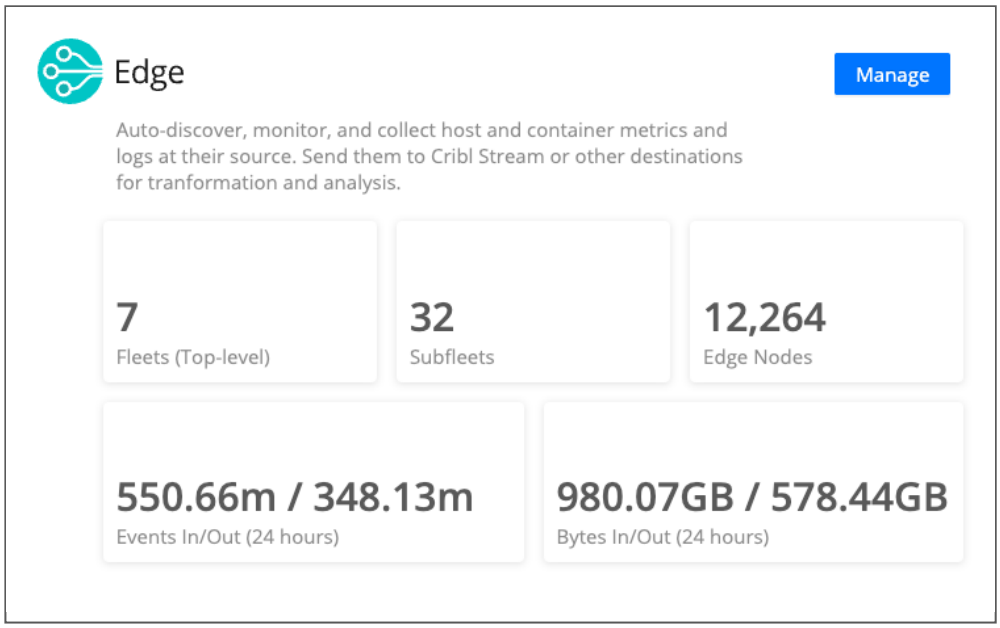
Functions process each event that passes through them. To help improve performance, Functions can optionally be configured with filters, to limit their processing scope to matching events only. For more details, see Functions.

1.3. Exploring Cribl Edge On Linux

The Cribl Edge UI offers a centralized view to manage, configure, and version–control your Edge Nodes. It also endows you with teleport–to–the–edge superpowers for locally previewing and validating your configurations. Here's a quick tour of the Cribl Edge UI in distributed mode.

Accessing Cribl Edge

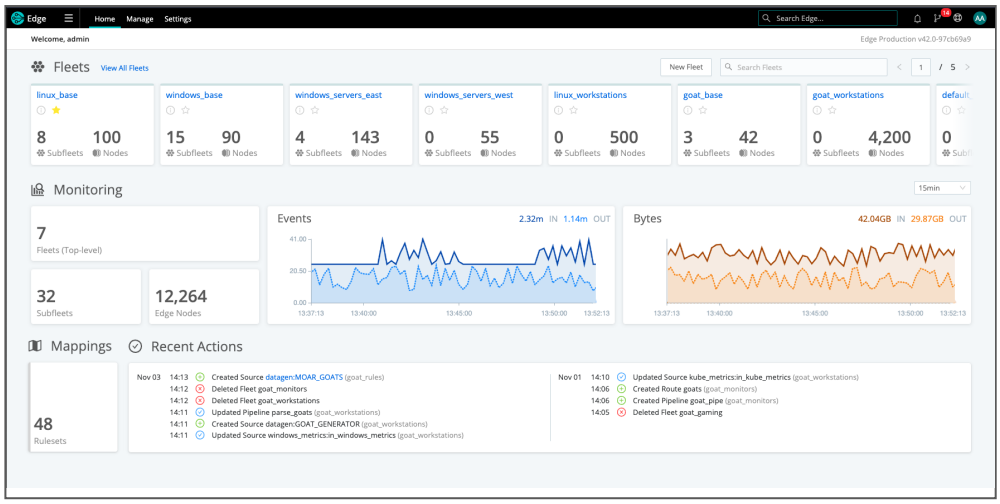
When you first log into Cribl Stream/Edge (single-instance or distributed), you'll see tiles that prompt you to choose between ~~two roads diverging in a yellow wood~~ the Stream versus Edge UIs. The Edge tile displays basic configuration details, including the number of Fleets, Subfleets, Edge Nodes, and events and bytes over time. Click **Manage** to start.



Manage Edge

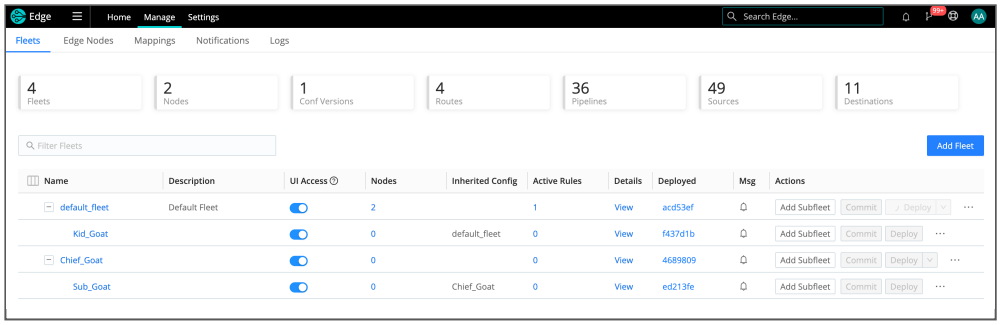
Fleets Overview

On the Cribl Edge **Home** tab, you can access your configured Fleets and a summary of your Cribl Edge environment, highlighting the aggregate data for all Fleets, Subfleets, Edge Nodes, and Mappings. The charts display information about traffic in and out of the system.



Cribl Edge Homepage

Select **Manage** from the top nav to view the **Fleets Landing** page. Here you can access the tabs for more information about your **Fleets** (and **Subfleets**), **Edge Nodes**, **Mappings**, Notifications, and **Logs**.

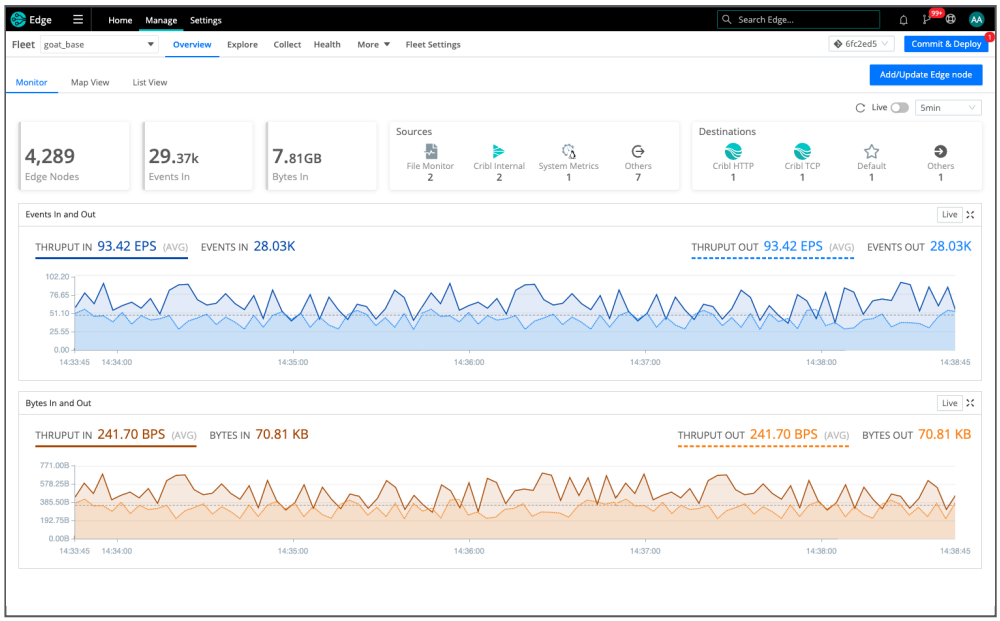


Manage Fleets

The **Manage Fleets** page gives you access to more information about your **Fleets** (and **Subfleets**), **Edge Nodes**, **Mappings**, Notifications, and **Logs**.

You can click a **Fleet** link to isolate individual Fleets, or use the **Search** bar to locate your Fleet.

Fleet Landing Page



Fleet Landing page

The Fleet's landing page highlights information about your configured Edge Nodes. The following information is displayed across the top.

Number of Edge Nodes: How many Edge Nodes are configured in this Fleet.

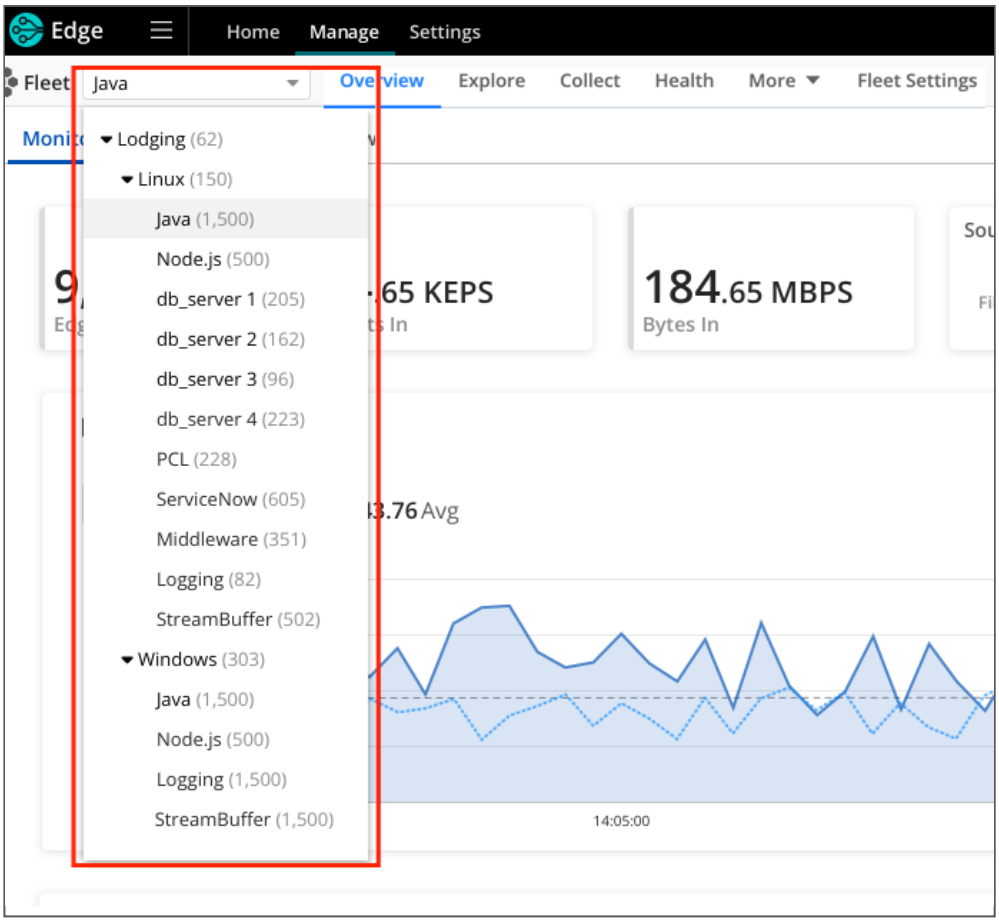
Events In: Total number of Events in the last 5 minutes of data collected. You can change the display's granularity from the default `last 5 min`, selecting a variety of time ranges from 1 min up to 1 day. (The latter covers the preceding 24 hours, and this maximum window is not configurable.)

Bytes In: The uncompressed amount of data in the last 5 minutes of data collected. You can change the display's granularity from the default `last 5 min`, selecting from a variety of time ranges from 1 min up to 1 day. (The latter covers the preceding 24 hours, and this maximum window is not configurable.)

Sources: List of configured Sources.

Destinations: List of configured Destinations.

Select the **Fleet** dropdown (top right) to see a hierarchical list of all your Fleets and Subfleets.



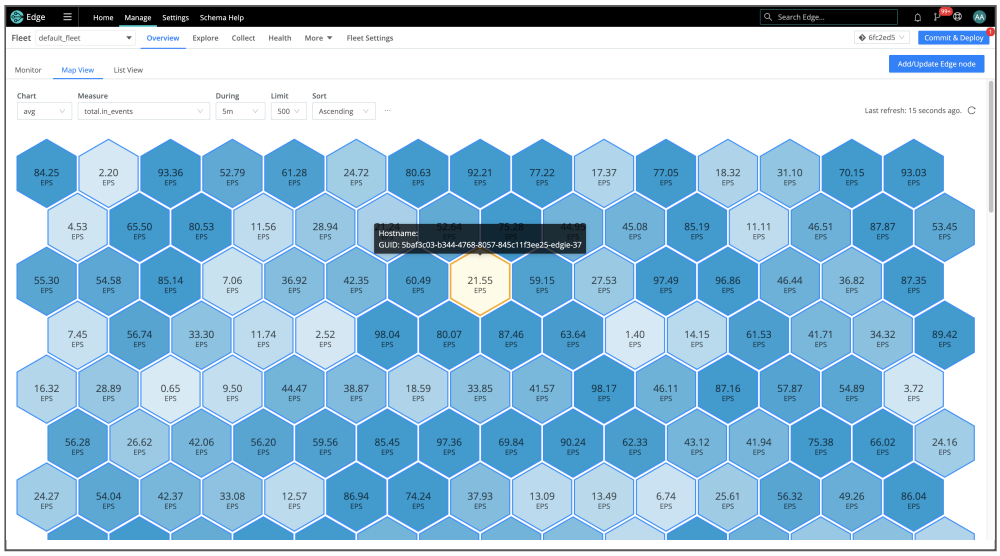
Fleets and SubFleets list

Fleet Map View

Map View: Here, a query builder allows you to display metrics from the Edge Nodes in the Fleet. You can select from different aggregations in the **Chart** field, different metrics in the **Measure** field, and the time window in the **During** field.

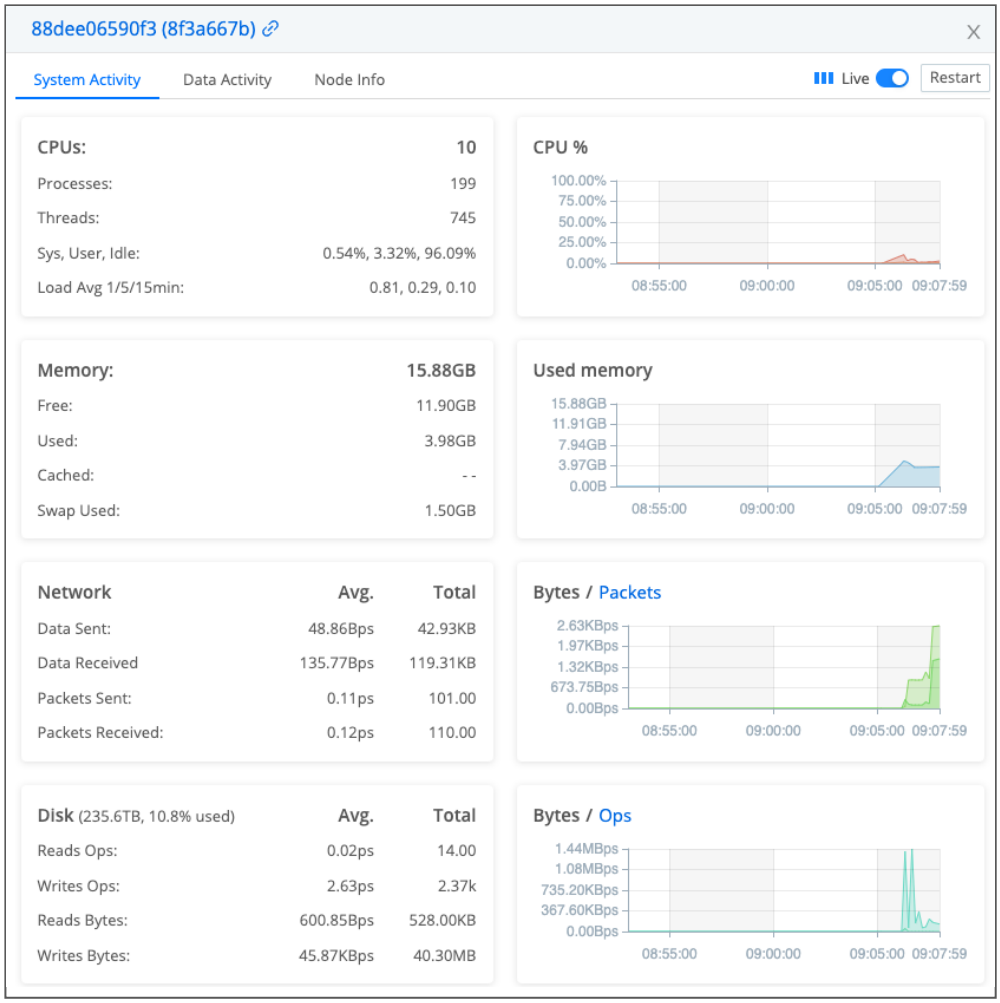
The metrics that appear in the **Measure** list depend on the option selected in **Fleet Settings** > **Limits** > **Metrics** under **Metrics to send from Edge Nodes**. The metrics `total.in_bytes`, `total.in_events`, `total.out_bytes`, and `total.out_events` always appear in the list; the display of any other metrics depends on what the Edge Nodes are sending to the Leader. To learn more about these metrics options, see Controlling Metrics.

A hexagon-based map view displays the resulting metric combination for each of your Edge Nodes. Hovering over on one of the hexagons displays the Edge Node's **GUID**.



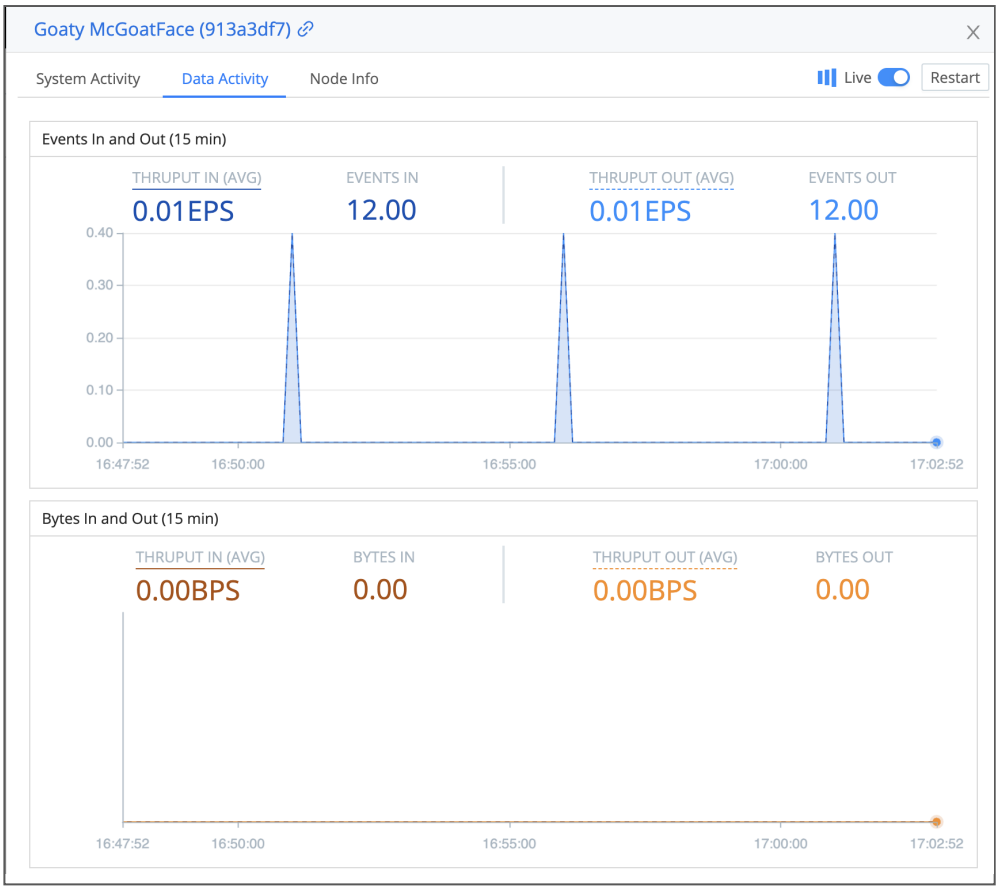
View Edge Node

Clicking any of the hexagons displays a modal providing details on the Edge Node, similar to teleporting into it, with the option to **Restart** the host. The **System Activity** tab displays details about the host's CPU, memory, network, and disk operations.



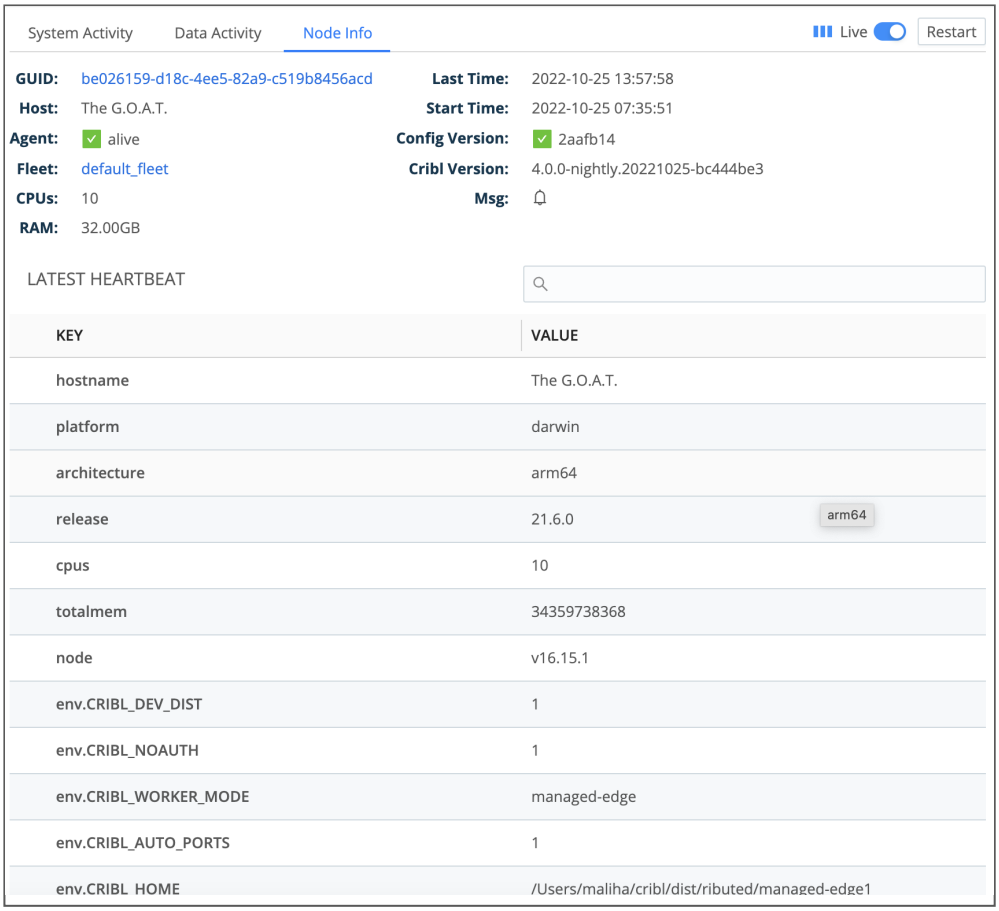
Edge Node System Activity

The **Data Activity** tab offers a view into the data flowing through the Edge Node.



Edge Node Data Activity

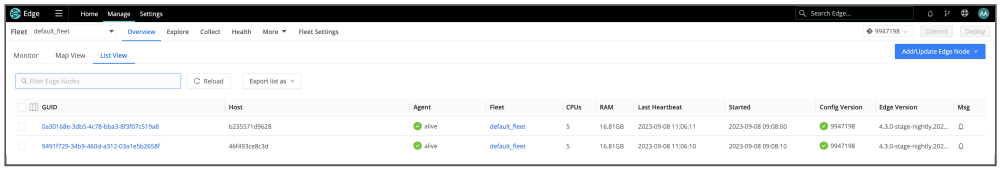
The **Node Info** tab offers Host/OS level information with a snapshot of the latest Heartbeat captured.



Edge Node info

Fleet List View

The **List View** tab displays a list of all the systems in the Fleet. This also serves as the “transporter room,” allowing you to teleport into each of the Node’s interfaces.



Edge Node list view

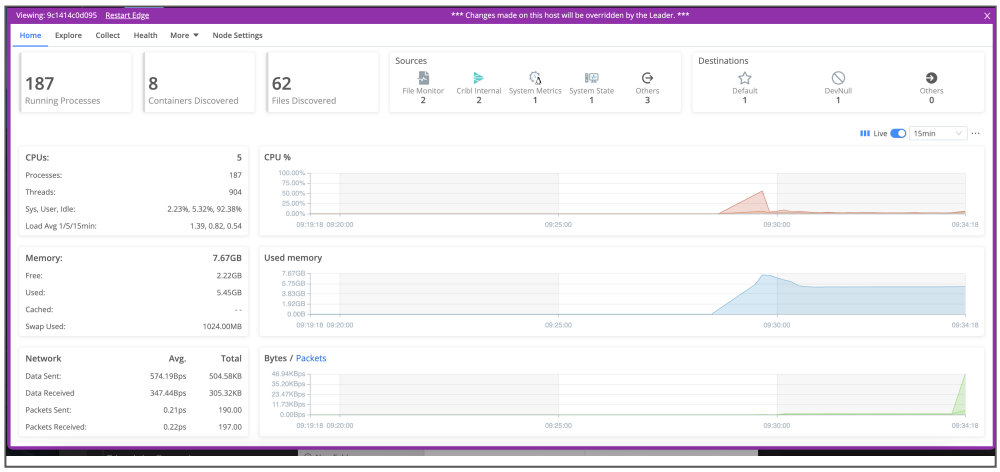
Click anywhere on a row to display a quick snapshot of the Edge Node’s details, with an option to **Restart** the host.

Teleport into an Edge Node

Click the **Edge Node GUID** link to teleport from the Leader into the Edge Node. Here, you can explore the metrics and log data that the Node has autodiscovered, and can manually discover and explore other data of interest. You can use the discovered data to perform root-cause analysis, to troubleshoot, and to restart the host.

The page displays metadata for the Node, and below it, graphs of system activity. A magenta border indicates you are remotely viewing a host, and identifies the host’s name.

Click **Restart Edge** to restart the Node. To return to the **Manage Edge Nodes** page, click the **X** close button on the upper right.



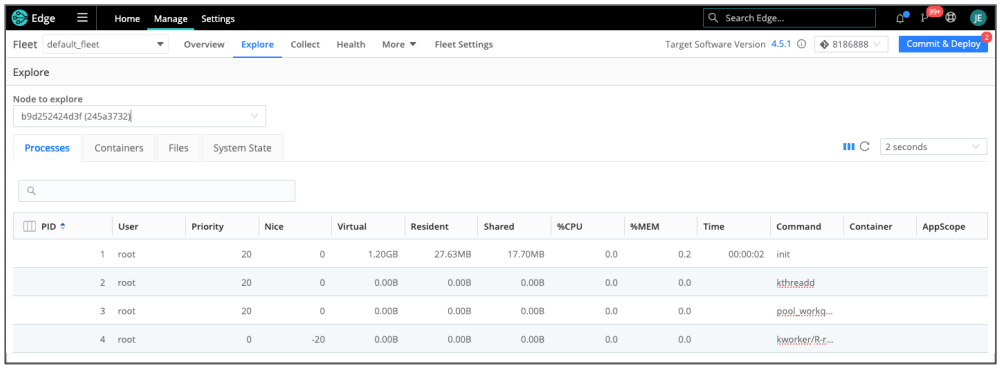
Teleporting into an Edge Node

Changes that you make on an Edge Node will not propagate to the Leader. Also, the Leader will override any changes that you make directly on a Node.

Explore Tab

From the top nav, select **Explore** to view more details on a particular Edge Node. On the **Node to explore** drop-down, select one of your hosts to display the following tabs:

- Processes
- Containers
- Files
- System State



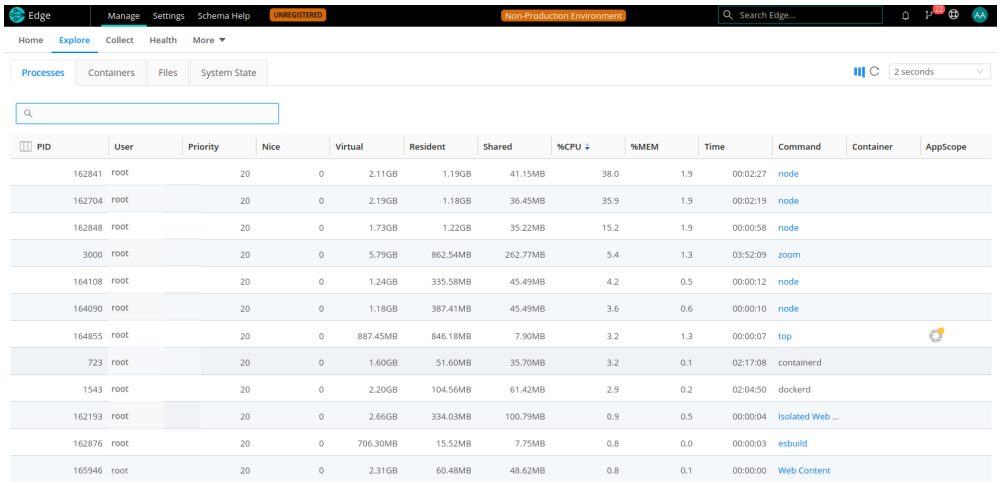
Explore an Edge Node

The **Node to explore** drop-down lists up to 50 Edge Nodes, ordered by hostname. To view the details for a specific Edge Node, enter the hostname or GUID into the **Node to explore** field.

Let's explore each tab.

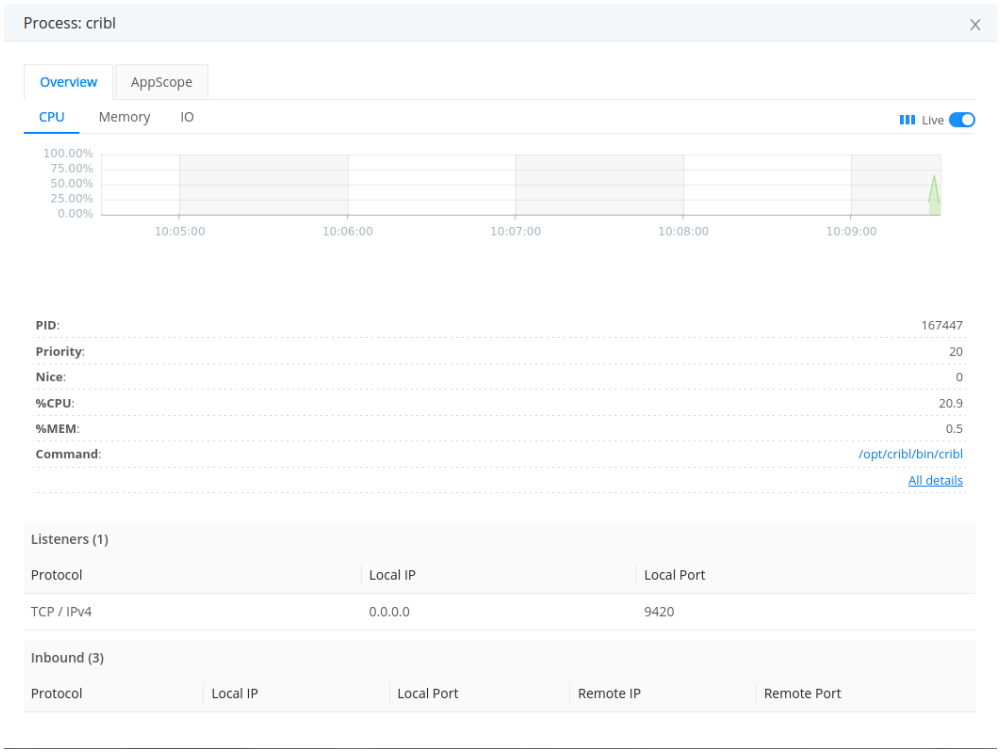
Processes

The **Processes** tab lists all the processes running on the Edge Node.



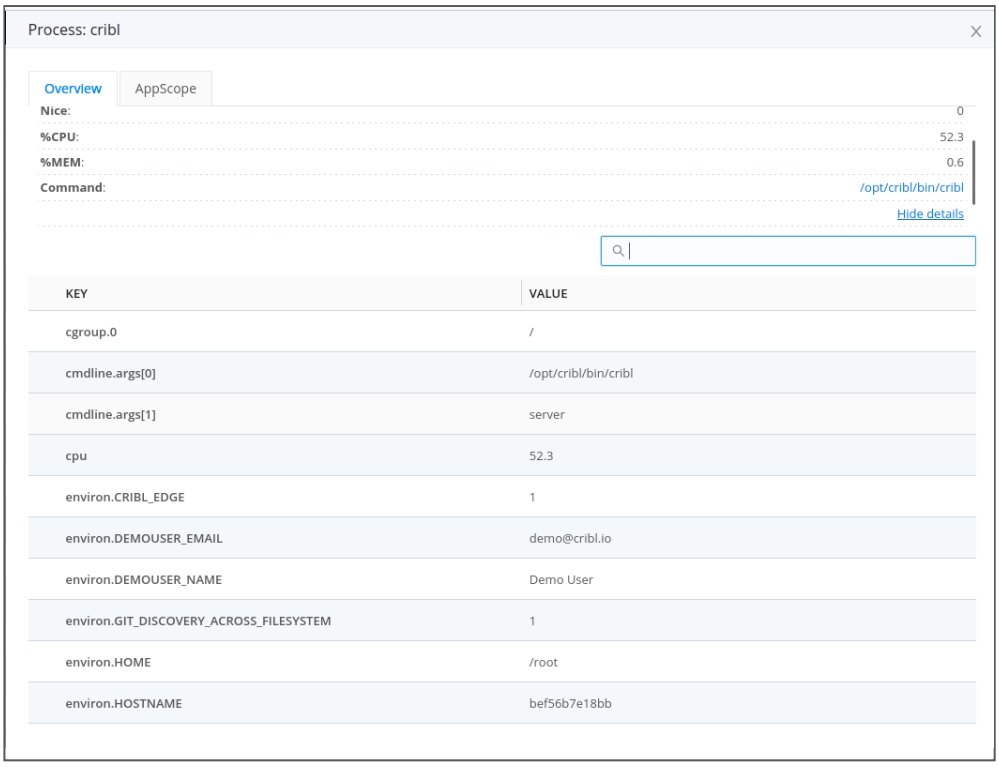
View Processes

Click on any of the rows to open the **Process: <process_name>** drawer. In the drawer's default **Overview** tab, you'll find basic information on the process, including CPU, Memory, and IO graphs, along with tables for active **Listening**, **Inbound**, and **Outbound** connections.



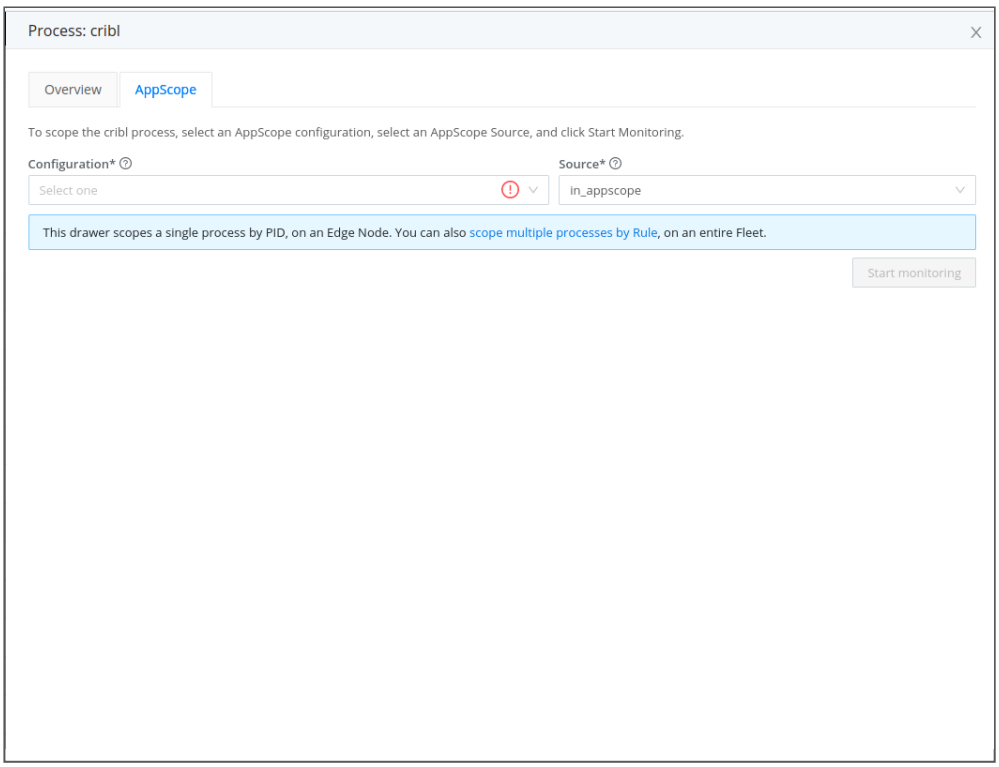
The Overview tab

In this tab, click **All details** to see the selected process' information out of `/proc`, expressed as key-value pairs. This information would normally require SSH'ing to the machine; this view makes troubleshooting across multiple systems much easier.



Showing all details of a process

Open the **AppScope** tab if you want to "scope" the process (i.e., use AppScope to monitor it). Once in the tab, you'll choose an an AppScope configuration that says what events and metrics to obtain, and an AppScope Source to receive them. See Scoping by PID in the AppScope docs. Note that your Cribl Edge instance must be running as root to do process monitoring with AppScope.



The AppScope tab for a process

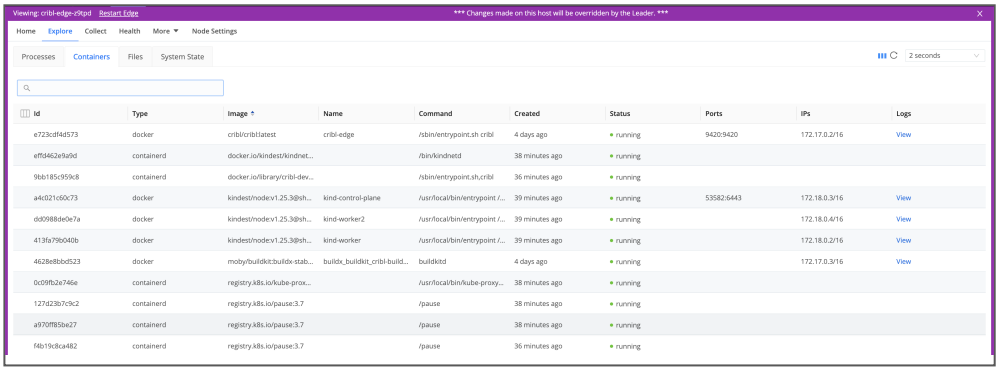
When a process is being scoped, back in the **Processes** tab you'll see that indicated in the process' entry in the **AppScope** column.

Containers

The **Containers** tab lists all the running containers and container metrics including information about images, volumes, status, ports, etc.

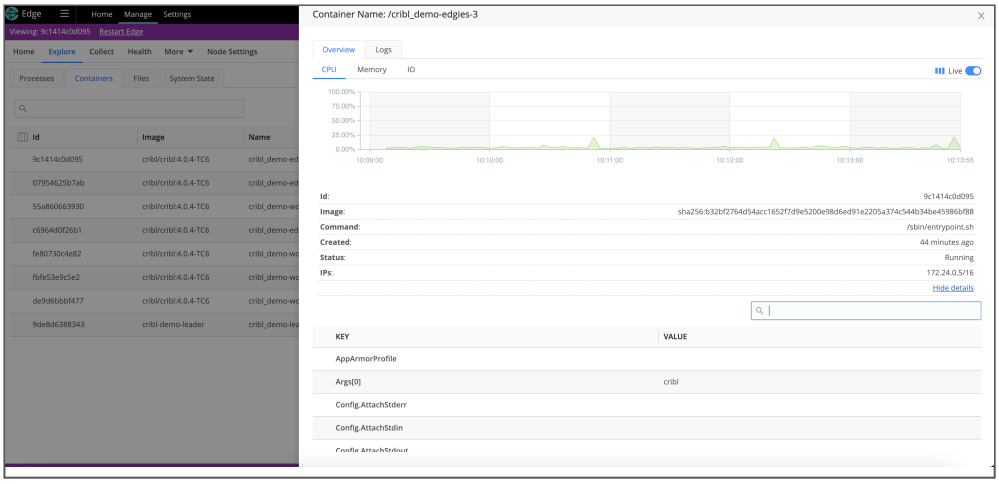
Cribl Edge supports both Docker and `containerd` runtimes.

`containerd` containers have less info than Docker containers, so `Ports`, `IPs`, and `Logs` won't populate.



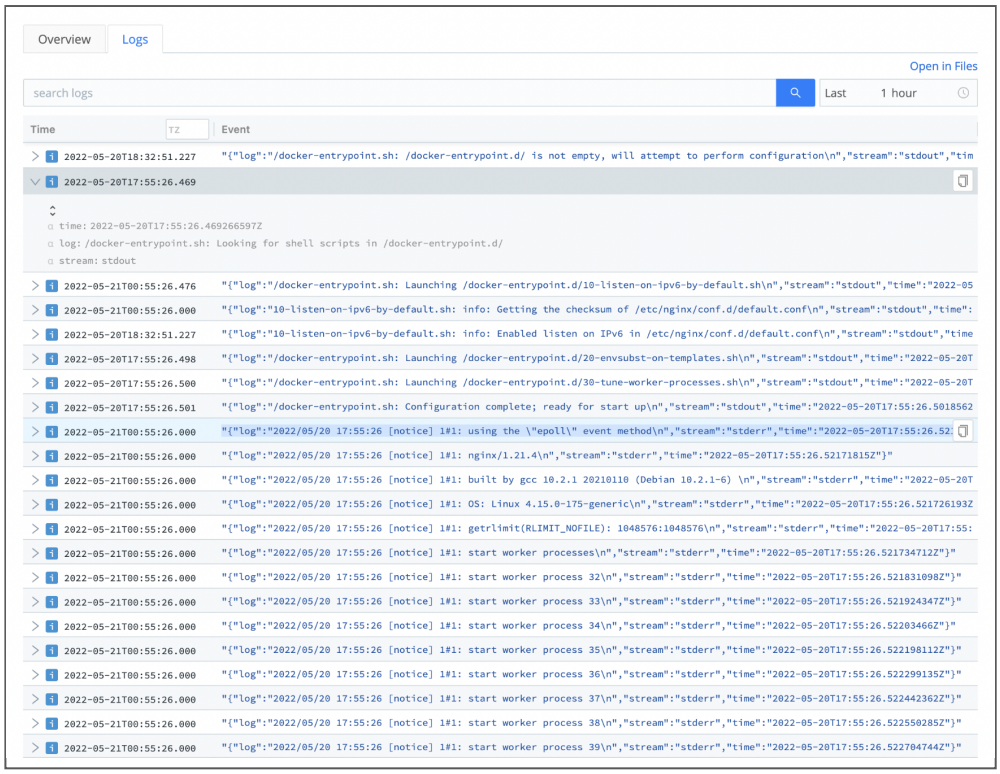
Containers overview

Click any container to view more details:



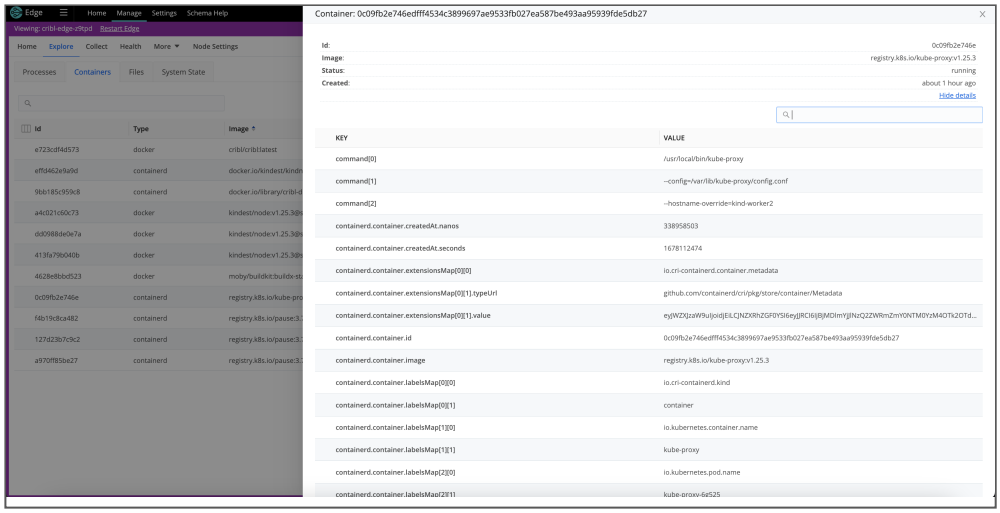
Container details

Click the **Logs** tab to view container logs. Optionally, use the search bar to filter displayed logs by arbitrary strings.



View Docker Container logs

The screenshot below shows `containerd` details, which don't include charts or logs.



View containerd logs

Files

The **Files** tab lists all the log files being actively written to by running applications that Cribl Edge has auto-discovered. You can also specify a list of directories and files to actively monitor.

Explore files

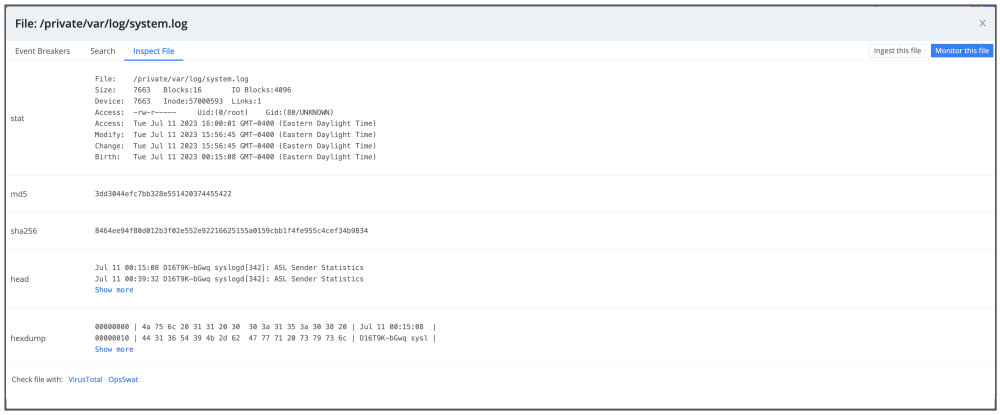
The **Actions** column allows you to:

- **View**: Displays a representation of the lines this column contains. You can also click any file row. To restrict how much data is displayed, use the search field or time picker on the **Search** tab.
- **Inspect**: Opens the **Inspect File** tab to show file metadata including details like permissions, file size, user, and modified date.

When inspected, compressed files (e.g. `foo.bar.gz`) include a **File preview** that shows the beginning of the file contents.

Archived files (e.g. `.zip`, `.tgz`, `.tar.gz`) include a **File listing** that shows the files within it.

If a file appears suspicious, click **VirusTotal** or **OpSwat** at the bottom of the modal to see if the file is flagged as compromised.



Inspect File tab

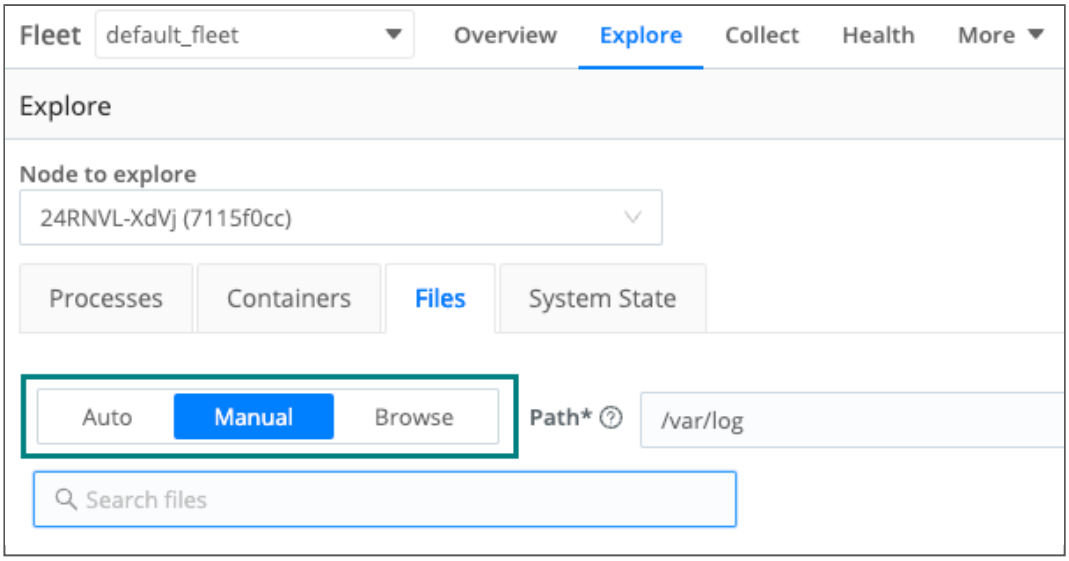
- **Monitor**: Displays the File Monitor's configuration modal.
- **Ingest**: Opens the Ingest file modal to send the file content to Routes/Pipelines for further processing or downstream to any destination you have configured. This is useful for testing and troubleshooting your configurations.

The **Files** tab provides the following options.

File Discovery Modes

Click a button at the top to select a discovery mode:

- **Auto**: Tells Cribl Edge to automatically discover files that are open for writing on currently running processes.
- **Manual**: Tells Cribl Edge to discover the files within the **Path** (directory) and **Allowlist** that you specify, down to the **Max depth**.
- **Browse**: Displays a tree view of all of your directories and files.



Manual discovery mode

Path

The **Path** field tells Cribl Edge to discover the files within the path (a directory) that you specify, down to the **Max depth**.

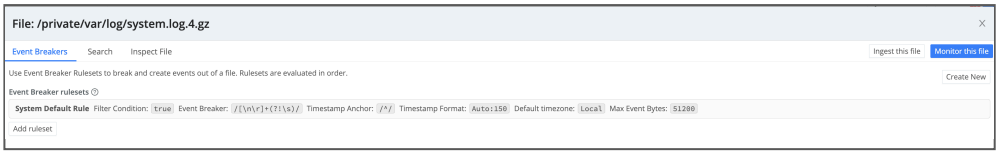
Allowlist

The **Allowlist** field, available with Auto and Manual discovery, supports wildcard syntax, and supports the exclamation mark (`!`) for negation. For example, you can use `!*cribl*access.log` to prevent Cribl Edge from discovering its own access log. The default filters are `*/log/*` and `*log`.

Click any file to see a representation of the lines it contains. To restrict how much data is displayed, you can use the search field or time picker on the **Search** tab.

Search tab

If the representation of events shown on the **Search** tab isn't ideally suited to the file's content, you can use the **Event Breakers** tab to change it.



Event Breakers tab

Monitor Files

The **Monitor Files** button, available with Auto and Manual discovery, opens a new File Monitor Source prefilled with the discovery mode and anything else you specified on the **Files** tab, such as allowlist entries, path, or max depth.

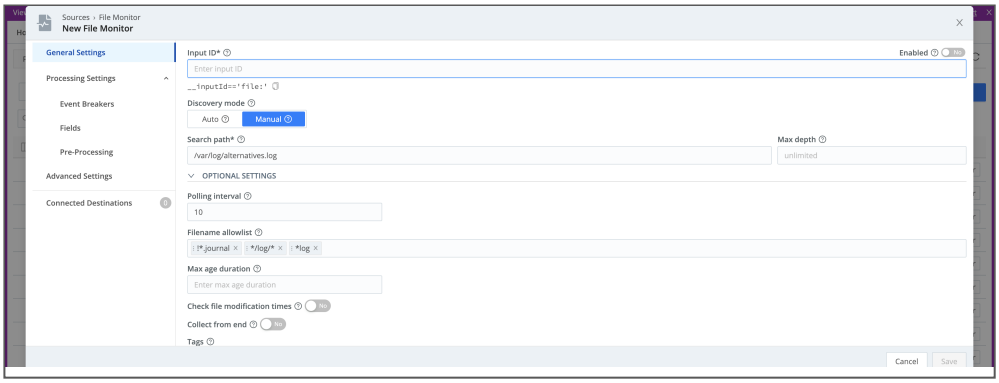
Max Depth

The **Max depth** field, available with Manual discovery, is empty by default. Cribl Edge will search subdirectories, and their subdirectories, downward without limit.

If you enter `0`, Cribl Edge will discover only the top-level files within the specified path. If you specify `1`, Cribl Edge will discover files one level down from the top. Follow this pattern to specify the depth you want.

Monitoring a File

Click a file's **Monitor** button or **Actions** option to configure your File Monitor Source to generate events from the file's lines or records.



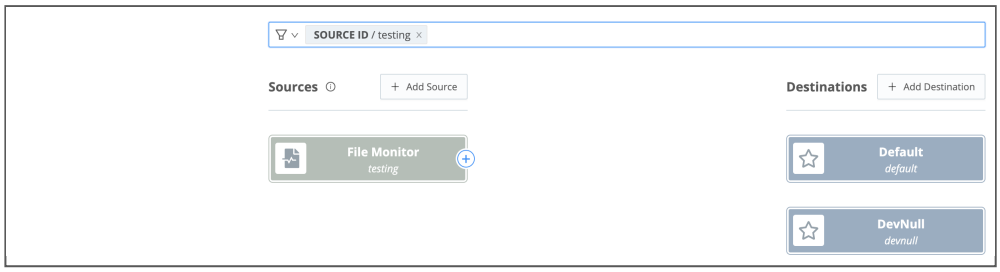
New File Monitor modal

The **Monitor** feature automatically prepopulates the modal with the following settings configured on the **Files** tab:

- **Discovery mode**
- **Search path**
- **Max depth**

- **Filename allowlist**

In addition, the **Connected Destinations** section defaults to **QuickConnect**. In the **Connected Destinations** section, you can select a Pipeline or Pack and a Destination. Otherwise, when you save, you'll be routed to the **Collect** page to set up your connections via QuickConnect.



Collect page

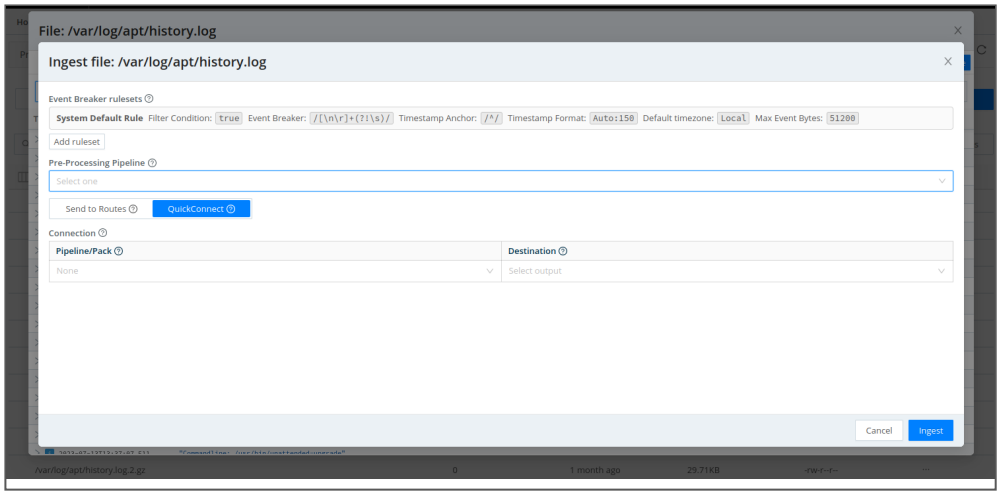
For further details, see File Monitor and QuickConnect.

Ingesting a File

To configure options for how and where to send file contents, use the **Ingest file** modal.

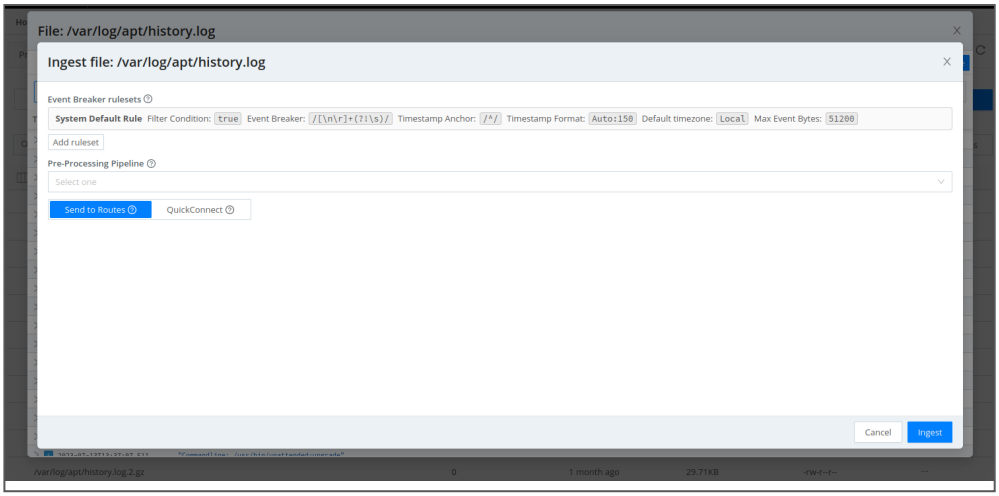
You have two options:

- Send directly to a configured Destination via **QuickConnect** (the default).



Ingest file via QuickConnect

- Send to Routes through (an optional) Pre-Processing Pipeline.



Ingest file via Routes

You can configure Event Breakers and rulesets for both options.

Exploring Files with Event Breakers

When you click a file in the **Files** tab, and Cribl Edge shows a representation of the lines that the file contains, how does that work? What's happening is that Cribl Edge is applying a default Event Breaker to the file.

You are not limited to the default Event Breaker, though. Select the **Event Breakers** tab, then:

- To apply a different (existing) Event Breaker, click **Add ruleset**, then select the desired ruleset from the **Event Breaker rulesets** drop-down.
- To create a new ruleset, click **Create New** to open the **New Ruleset** modal. Proceed as described here. Later, you can persist the new Event Breaker as part of a Source or a Collector. While you create the new ruleset, Cribl Edge pulls the contents of the open file into the Sample File area. Toggle between the **In** and **Out** tabs to compare, respectively, the original content, and the content as modified by the Event Breaker you're creating.

Now return to the **Search** tab – the contents of your chosen file will appear with the new Event Breaker applied.

System State

The **System State** upper tab provides access to these left tabs:

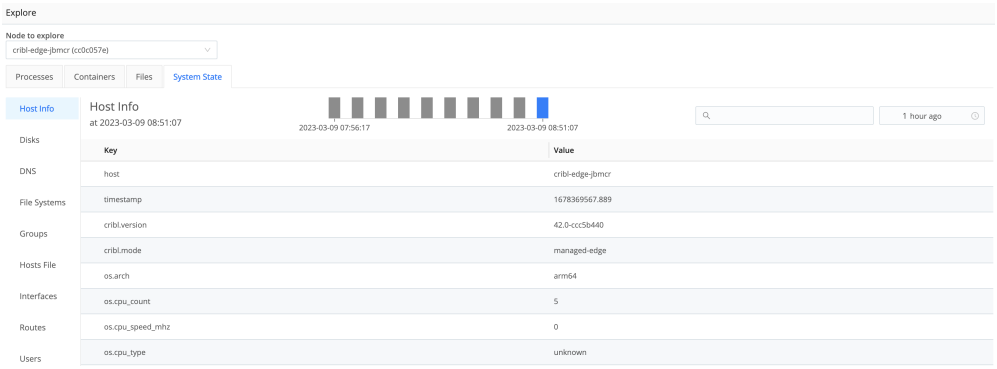
- Host Info
- Disks
- DNS
- File Systems
- Firewall

- Groups
- Hosts File
- Interfaces
- Listening Ports
- Logged-In Users
- Routes
- Services
- Users

To display any of the tabs above, you need to configure and enable the System State Source. Also, make sure that the Source's Collector Settings fields are enabled.

Host Info

Cribl Edge can add a `__metadata` property to every event emitted from every enabled Source. The **System State** tab displays the metadata collected for each Edge Node under **Host Info**.



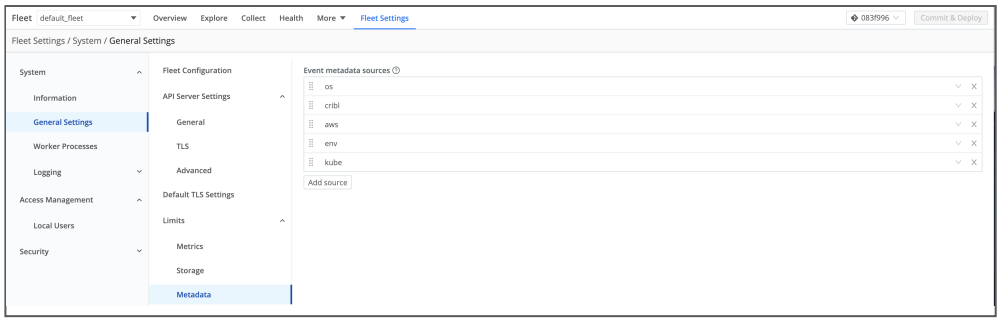
Host info/metadata collected

The metadata surfaced by an Edge Node can be used to:

- Enrich events (with an internal `__metadata` field).
- Display to users as a part of instance exploration.

You can customize the type of metadata collected at **Fleet Settings** > **Limits** > **Metadata**. Use the **Event metadata sources** drop-down (and/or the **Add source** button) to add and select metadata sources.

In Edge mode, all the **Event metadata sources** are enabled by default.



Set limits

The metadata sources that you can select here include:

- `os`: Reports details for the host OS and host machine, like OS version, kernel version, CPU and memory resources, hostname, network addresses, etc.
- `cribl`: Reports the Cribl Edge version, mode, Fleet for managed instances, and config version.
- `aws`: Reports details for an EC2 instance, including the instance type, hostname, network addresses, tags, and IAM roles. For security reasons, we report only IAM role names.
- `env`: Reports environment variables.
- `kube`: Reports details on a Kubernetes environment, including the node, Pod, and container. For details, see __metadata.kube Property.

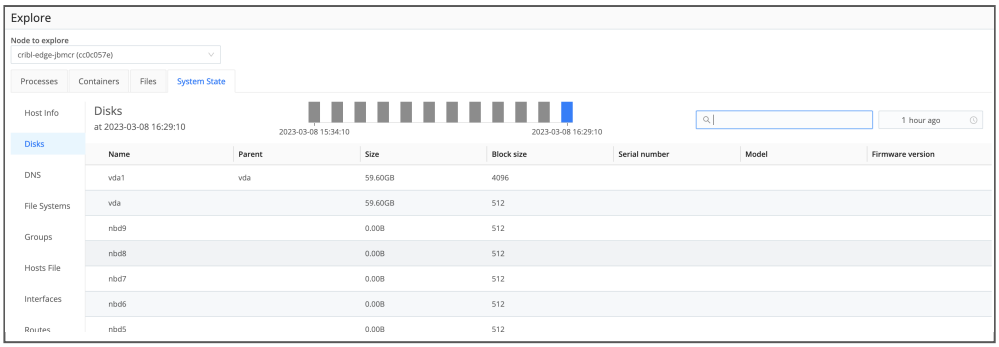
When these metadata sources are enabled (and can get data), Cribl Edge will add the corresponding property to events, with a nested property for each enabled source.

Some metadata sources work only in configured environments. For example, the `aws` source is available only when running on an AWS EC2 instance.

If your security tools report denied outbound traffic to IP addresses like `169.254.169.168` or `169.254.169.254`, you can suppress these by removing `aws` from the metadata sources described above. If you have a proxy setup, Cribl recommends adding these IP addresses to your `no_proxy` environment variable.

Disks

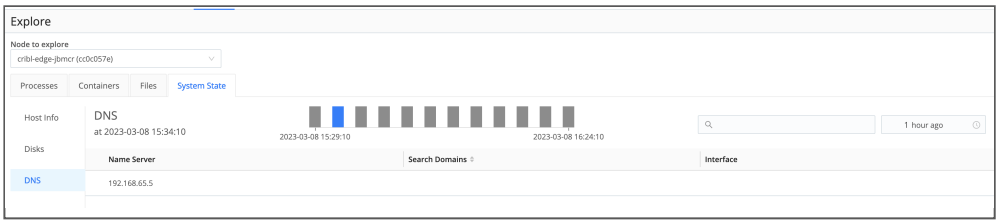
The **Disks** tab displays the inventory of physical disks and their partitions on the host system.



Disks tab

DNS

The **DNS** tab lists the host system's DNS resolvers and search entries.



DNS tab

File Systems

The **File Systems** tab displays an inventory of the mounted file systems on the host system.

File Systems tab

Firewall

The **Firewall** tab displays a list of the host's defined firewall rules.

Firewall tab

Groups

The **Groups** tab displays a list of the local groups including their names, descriptions, and members on the host system.

Groups tab

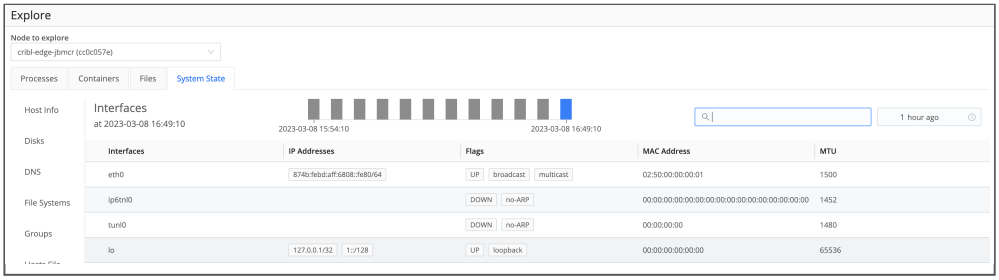
Hosts File

The **Hosts File** tab displays the host system's current state.

Hosts File tab

Interfaces

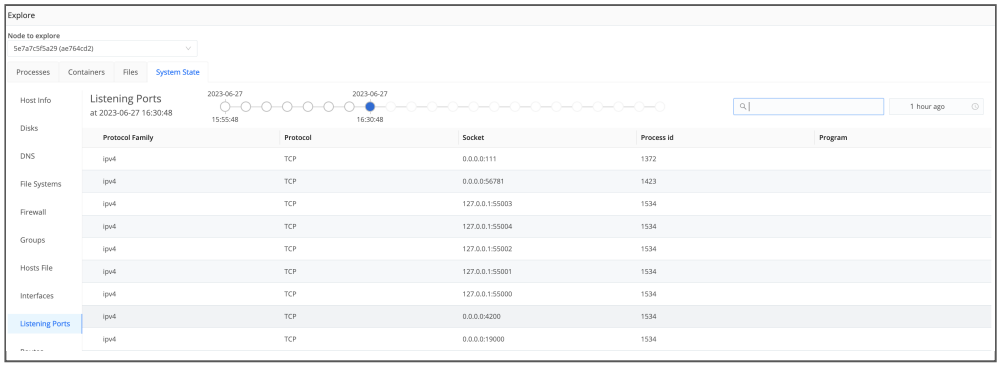
The **Interfaces** tab displays a list of each of the network interfaces on the host system.



Interfaces tab

Listening Ports

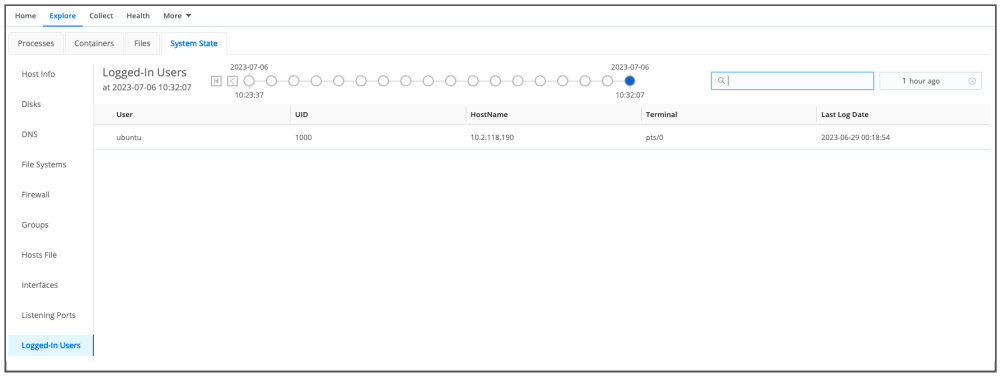
The **Listening Ports** tab displays a list of listening ports and their associated process identifier (pid).



Listening Ports tab

Logged-In Users

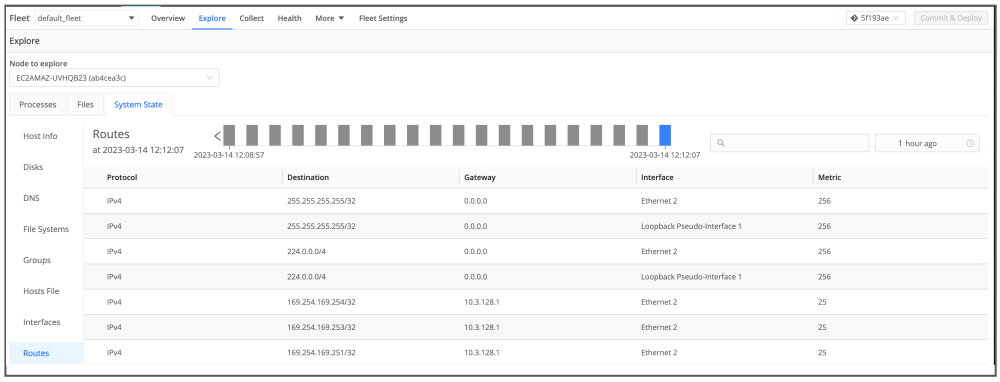
The **Logged-In Users** tab displays a list of currently logged-in users on the host.



Logged-In Users tab

Routes

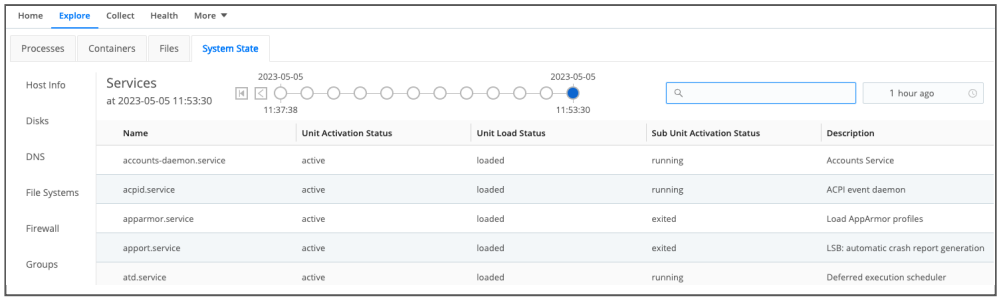
The **Routes** tab displays entries from the network routes on the host system.



Hosts Routes tab

Services

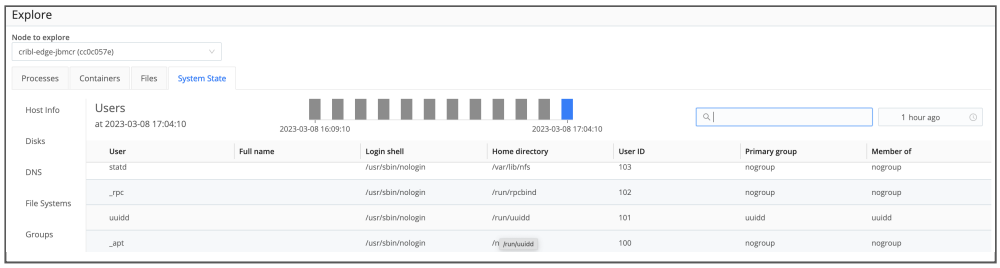
The **Services** tab displays a list of each configured service (e.g. `systemd` and `initd`) along with their running status.



Services tab

Users

The **Users** tab displays a list of local users on the host system.



Users tab

Cribl University offers a course titled Collecting Data with Linux that provides a good overview of working with Linux. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

`__metadata.kube` Property

For the `__metadata.kube` property (`kube` in the UI) to report details on a Kubernetes environment, Cribl Edge needs to figure out where it is running. Set the `KUBE_K8S_POD` environment variable to the name of the Pod in which Cribl Edge is running. At this point, the `__metadata.kube` property will have information to report on the `node` and `pod` properties.

If the `/proc/self/cgroup` is working, then the `container` property information will be available, too.

If you leave the `KUBE_K8S_POD` environment variable unset, and `/proc/self/cgroup` is not working, then Cribl Edge will not know what Pod it is running in. This state has multiple implications:

- The Kubernetes Metrics Source will be unable to identify whether or not it is in a DaemonSet. The result will be redundant metrics from each node in the cluster.
- The Kubernetes Metadata collector will not add the `__metadata.kube property`.
- The Kubernetes Logs Source will also duplicate data. Every node in the cluster will collect logs for every container in the cluster.

1.4. Exploring Cribl Edge On Windows

Cribl Edge on Windows offers easy-to-use tools for exploring and collecting Windows events. You can run Cribl Edge on Windows Server 2016, 2019, or 2022, to collect events via the Windows Events API.

Limitations

Cribl Edge on Windows is currently subject to the following limitations:

Modes

Cribl Edge on Windows supports only the following modes:

- `Edge: Single`
- `Edge: Managed Edge (managed by Leader)`

This means you can't switch Cribl Edge on Windows into Cribl Stream mode (Single-instance, Worker, or Leader).

You can, however, switch between the Cribl Edge supported modes via the UI, in **Settings** (top nav) > **Distributed Settings** > **Mode**.

Do not select an unsupported mode from this drop-down! Doing so will cause the Cribl service to fail.

Sources and Destinations

Cribl Edge on Windows supports the same Sources and Destinations as Cribl Edge on Linux, with the following exceptions:

- The AppScope and System Metrics Sources are unavailable on Windows.
- The Kubernetes Logs and Kubernetes Metrics Sources. Cribl Edge on Windows does not support Kubernetes deployments.
- The Kafka Destination does not support Kerberos authentication on Windows.

Sources on Windows Only

The following Sources are available **only** when running Cribl Edge on Windows (not on Linux):

- Windows Event Logs
- Window Metrics

Functions

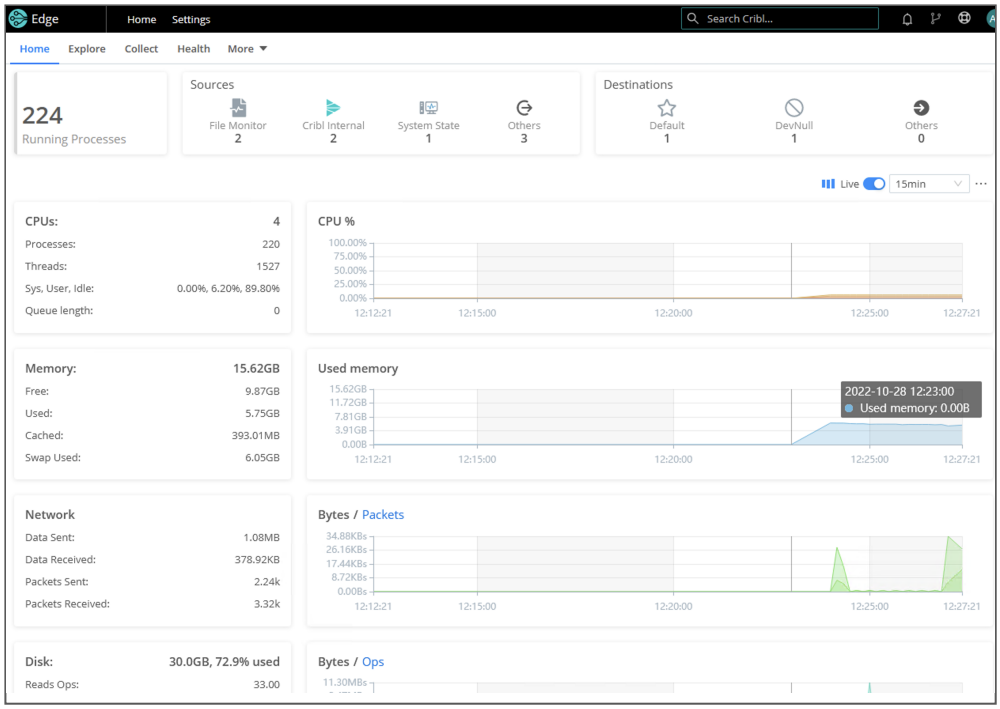
The Grok Function is unavailable on Windows. If you include it in a Pipeline, Cribl Edge processing will skip over it.

Data Formats

Cribl Edge on Windows does not support reading or writing Parquet files. This is a limitation on the following Destinations: Amazon S3, Azure Blob Storage, MinIO, and FileSystem/NFS.

Accessing Cribl Edge on Windows

When you first log into Cribl Edge on Windows (single-instance or managed node), you'll land on the **Home** tab where you can explore the metrics and log data that the Node has auto-discovered, and can manually discover and explore other data of interest.



Edge on Windows

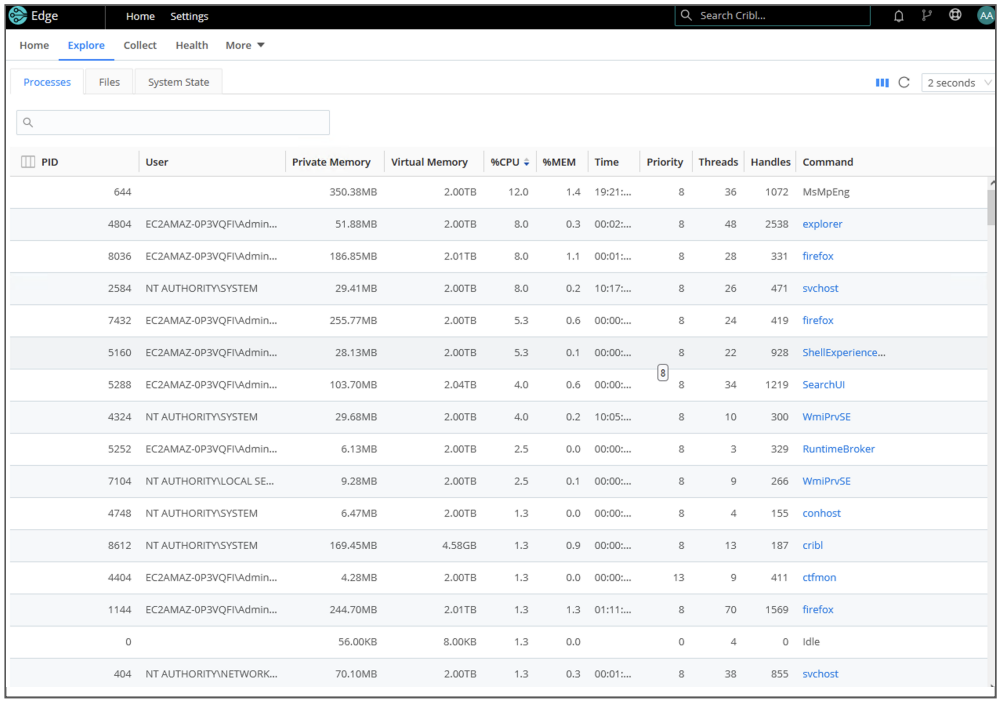
From the **Explore** page, you can view more details on your node via the following tabs:

- Processes
- Files

- Host Info/Metadata

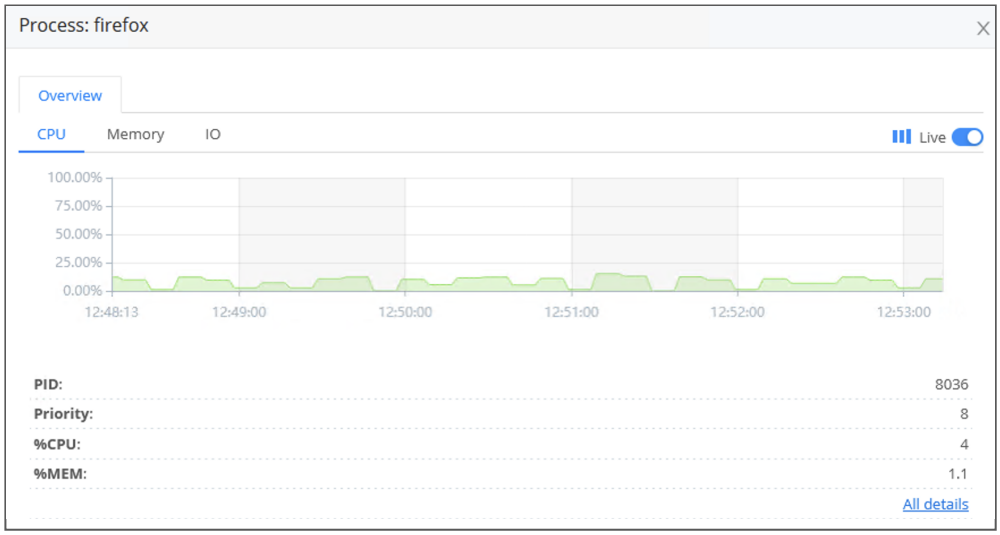
Processes

The **Processes** tab lists all the processes running on the Edge Node.



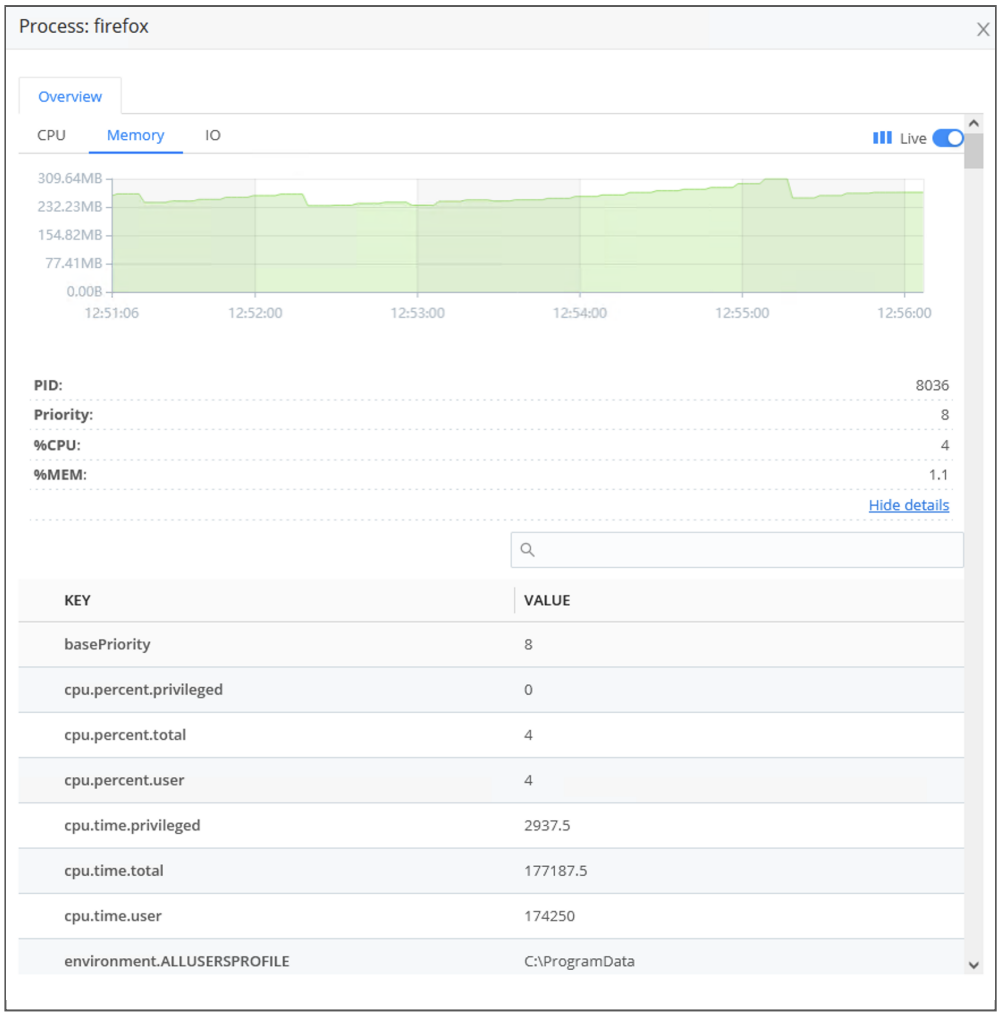
Processes tab

Click on any of the rows to display a modal with basic information on the process, including CPU usage, Memory usage, and I/O graphs.



Process details

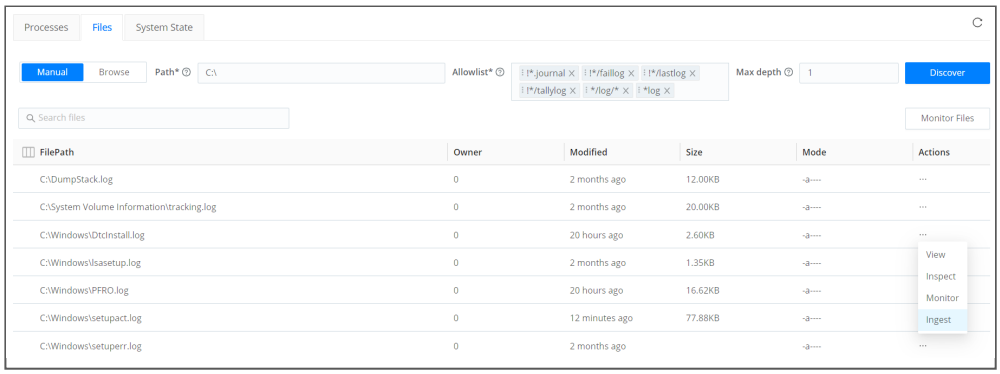
In this modal, click **All details** to get a table view of processes' information. To access this information, you would normally need to SSH to the machine; this view makes troubleshooting across multiple systems much easier.



All detail link

Files

The **Files** tab enables you to specify a list of directories and files to actively monitor.

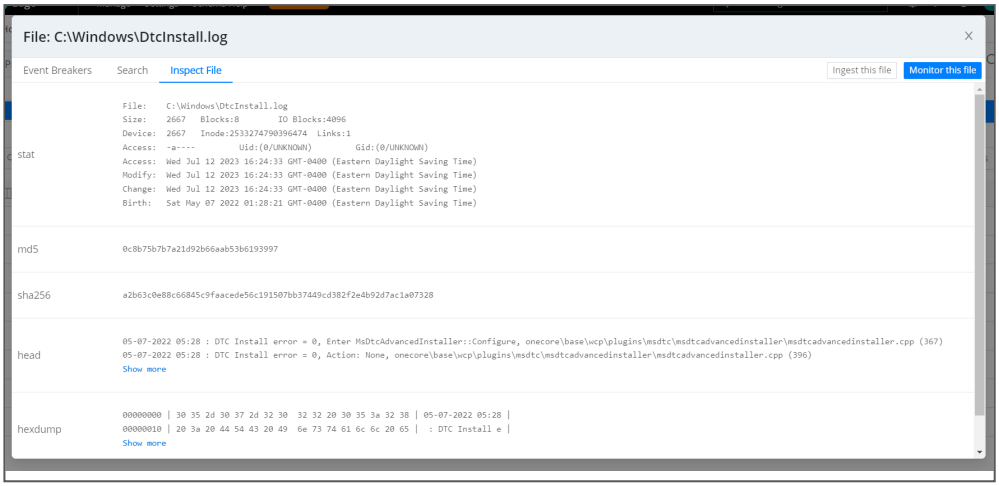


File tab -Manual mode

The **Actions** column allows you to:

- **View**: Displays a representation of the lines this column contains. You can also click any file row. To restrict how much data is displayed, use the search field or time picker on the **Search** tab.

- **Inspect**: Opens the **Inspect File** tab to show file metadata, including details like permissions, file size, user, and modified date.



Inspect File tab

- **Monitor**: Displays the File Monitor's configuration modal.
- **Ingest**: Opens the Ingest file modal to send the file content to Routes/Pipelines for further processing or downstream to any destination you have configured. This is useful for testing and troubleshooting your configurations.

The **Files** tab provides the following options.

File Discovery Modes

There are two discovery modes:

- **Manual**
- **Browse**

The **Manual** mode provides the following options:

Path

The **Path** field tells Cribl Edge to discover the files within the path (a directory) that you specify, down to the **Max depth**.

Allowlist

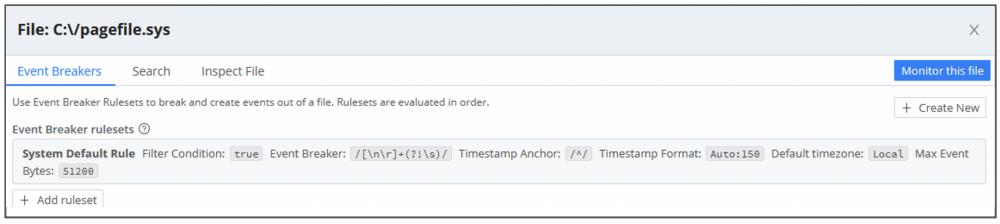
The **Allowlist** field supports wildcard syntax, and supports the exclamation mark (!) for negation.
For example, you can use `!*cribl*access.log` to prevent Cribl Edge from discovering its own access log.

The default filters are `*/log/*` and `*log`.

Click any file to see a representation of the lines it contains. To restrict how much data is displayed, you can use the search field or time picker on the **Search** tab.

Search tab

If the representation of events shown on the **Search** tab isn't ideally suited to the file's contents, you can use the **Event Breakers** tab to refine it.



Event Breakers tab

Max Depth

The **Max depth** field is empty by default. Cribl Edge will search subdirectories, and their subdirectories, downward without limit.

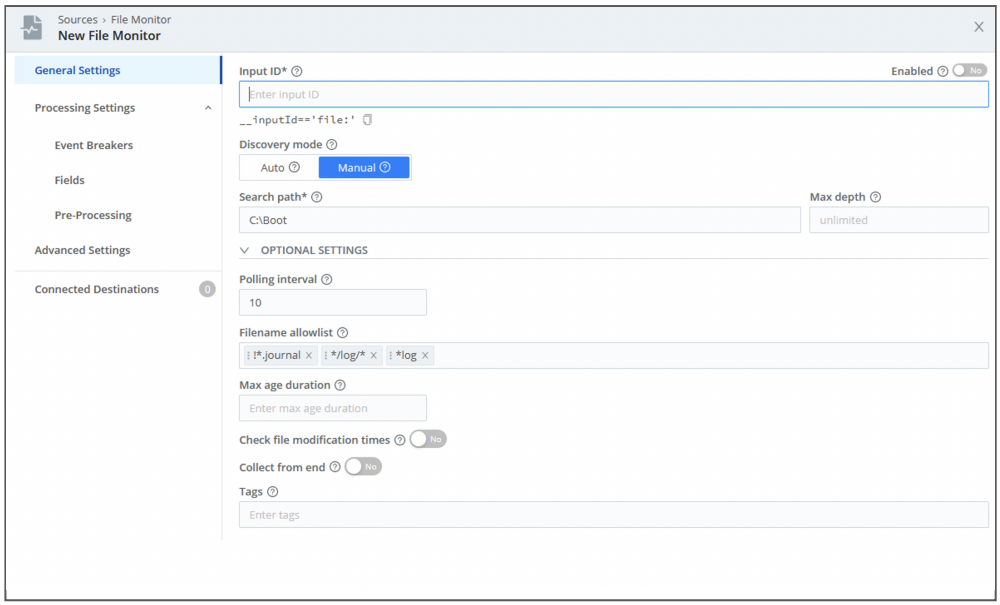
If you enter `0`, Cribl Edge will discover only the top-level files within the specified path. If you specify `1`, Cribl Edge will discover files one level down from the top. Follow this pattern to specify the depth you want.

Monitor Files

The **Monitor Files** button opens a **New File Monitor** modal prefilled with the discovery mode, path, max depth, and allowlist entries specified on the **Files** tab.

Monitor

Click a file's **Monitor** button to configure your File Monitor Source to generate events from the file's lines or records.

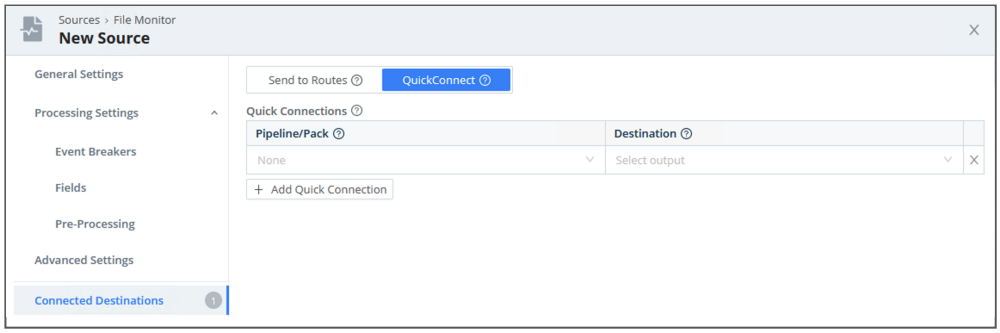


New File Monitor modal

The **Monitor** feature automatically prepopulates the modal with the following settings configured on the **Files** tab:

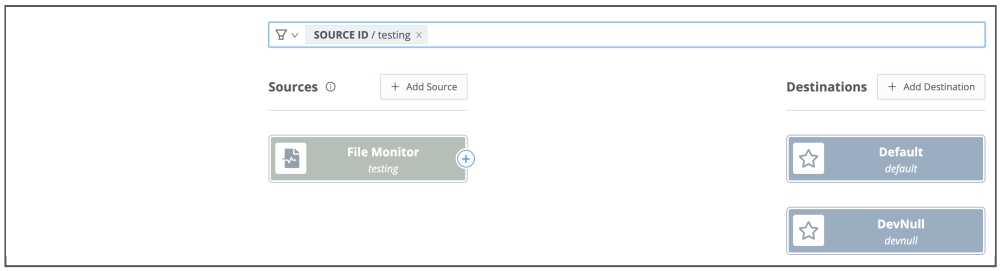
- **Search path**
- **Discovery mode**: Defaults to **Manual** for Windows.
- **Max depth**
- **Filename allowlist**

Note that the **Connected Destinations** section defaults to **QuickConnect**, Cribl Edge's graphical UI. In the **Connected Destinations** section, you can select a Pipeline or Pack and a Destination. Otherwise, when you save, you'll be routed to the **Collect** page to set up your connections via QuickConnect.



Connected Destinations

For details about making these connections, see File Monitor and QuickConnect.



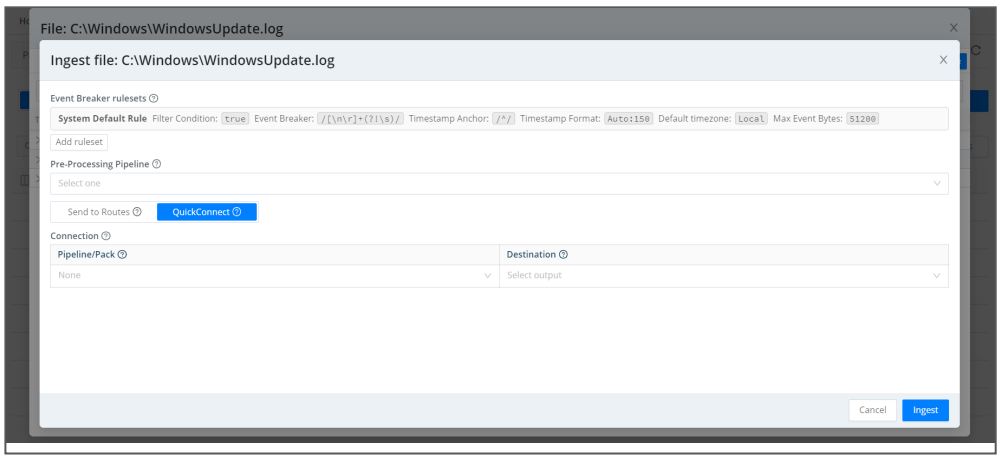
Collect page

Ingesting a File

To configure options for how and where to send file contents, use the **Ingest file** modal.

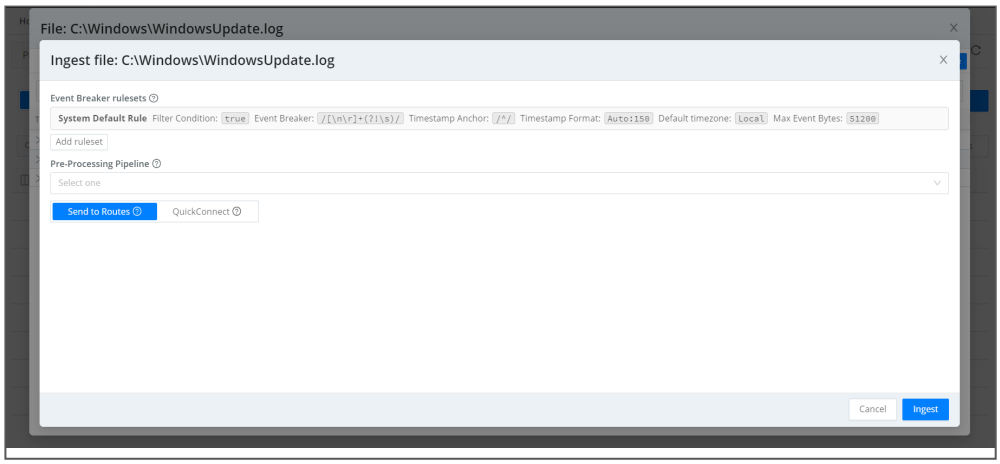
You have two options:

- Send directly to a configured Destination via **QuickConnect** (the default).



Ingest file via QuickConnect

- Send to Routes through (an optional) Pre-Processing Pipeline.



Ingest file via Routes

You can configure Event Breakers and rulesets for both options.

Exploring Files with Event Breakers

When you click a file in the **Files** tab, and Cribl Edge shows a representation of the lines that the file contains – how does that work? Cribl Edge is applying a default Event Breaker to format the file.

You are not limited to the default Event Breaker, though. Select the **Event Breakers** tab, then:

- To apply a different (existing) Event Breaker, click **Add ruleset**, then select the desired ruleset from the **Event Breaker rulesets** drop-down.
- To create a new Event Breaker ruleset, click **Create New**. In the resulting **New Ruleset** modal, proceed as described here. Later, you can reuse the new Event Breaker as part of a Source or a Collector. While you create the new ruleset, Cribl Edge pulls the contents of the open file into the Sample File area. Toggle between the **In** and **Out** tabs to compare, respectively, the original content and the content as modified by the Event Breaker you're creating.

Now return to the **Search** tab – the contents of your chosen file will appear with the new Event Breaker applied.

System State

The **System State** upper tab provides access to these left tabs:

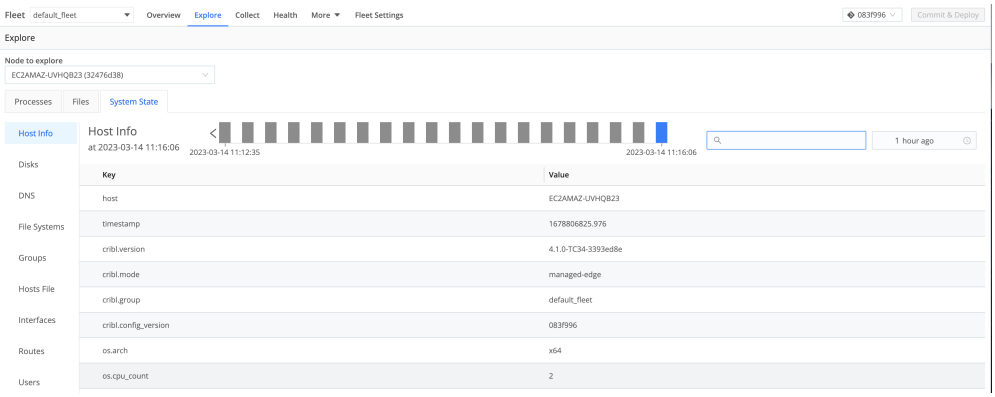
- Host Info
- Disks
- DNS
- File Systems
- Firewall
- Groups
- Hosts File
- Interfaces
- Listening Ports
- Logged-In Users
- Routes
- Services
- Users

To display any of the tabs above, you need to configure and enable the System State Source. Also,

make sure that the Source's Collector Settings fields are enabled.

Host Info/Metadata

Cribl Edge can add a `__metadata` property to every event emitted from every enabled Source. The **Host Info** tab displays the metadata collected for each Edge Node.



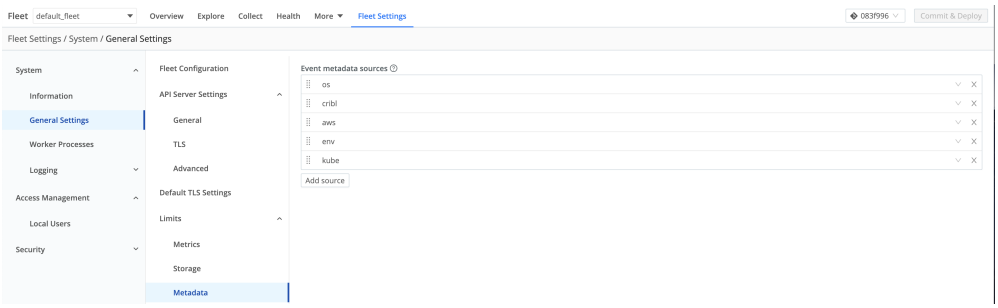
Host info / Metadata collected

The metadata surfaced by an Edge Node can be used to:

- Enrich events (with an internal `__metadata` field).
- Display to users as a part of instance exploration.

To customize the type of metadata collected, select **Settings** > **General Settings** > **Limits** > **Metadata**. Use the **Event metadata sources** drop-down (and/or the **Add source** button) to add and select metadata sources.

In Edge mode, all the **Event metadata sources** are enabled by default.



Set limits

The metadata sources that you can select here include:

- `os`: Reports details for the host OS and host machine, like OS version, kernel version, CPU and memory resources, hostname, network addresses, etc.
- `cribl`: Reports the Cribl Edge version, mode, Fleet for managed instances, tags defined on the instance, and config version.
- `aws`: Reports details for an EC2 instance, including the instance type, hostname, network addresses, tags, and IAM roles. For security reasons, we report only IAM role names.
- `env`: Reports environment variables.

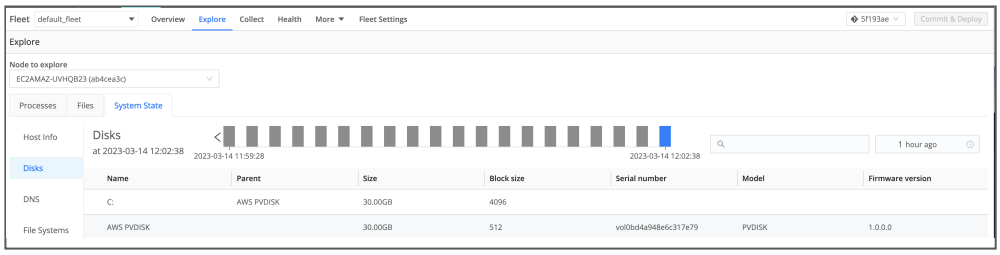
When these metadata sources are enabled (and can get data), Cribl Edge will add the corresponding property to events, with a nested property for each enabled source.

Some metadata sources work only in configured environments. For example, the `aws` source is available only when running on an AWS EC2 instance.

If your security tools report denied outbound traffic to IP addresses like `169.254.169.168` or `169.254.169.254`, you can suppress these by removing `aws` from the metadata sources described above. If you have a proxy setup, Cribl recommends adding these IP addresses to your `no_proxy` environment variable.

Disks

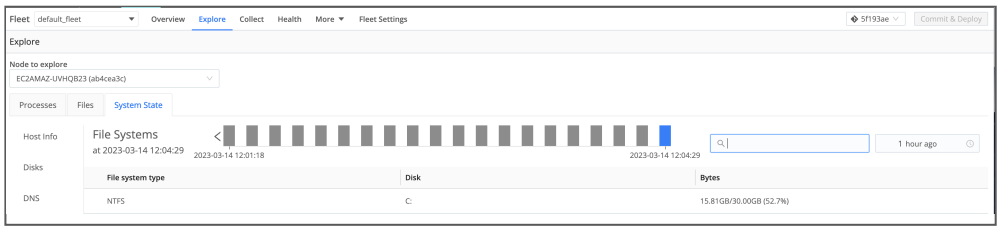
The **Disks** tab displays the inventory of physical disks and their partitions on the host system.



Disks tab

DNS

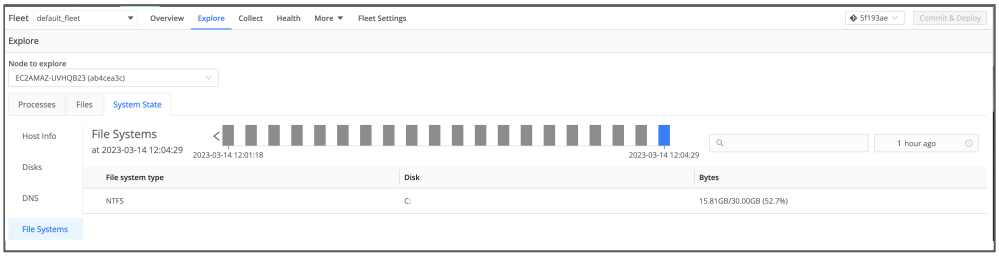
The **DNS** tab lists the DNS resolvers and search entries on the host system.



DNS tab

File Systems

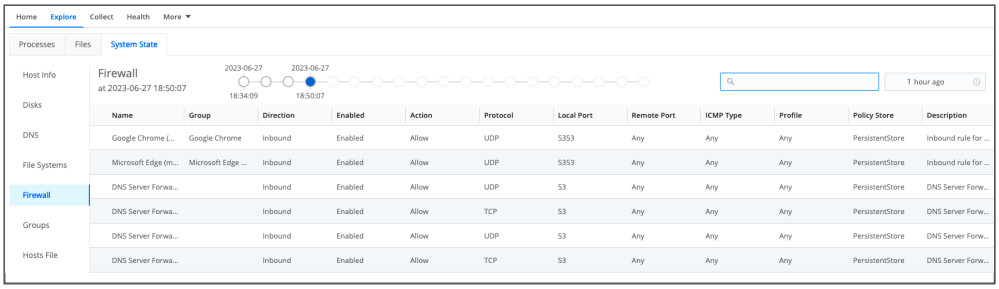
The **File Systems** tab displays an inventory of the mounted file systems on the host system.



File Systems tab

Firewall

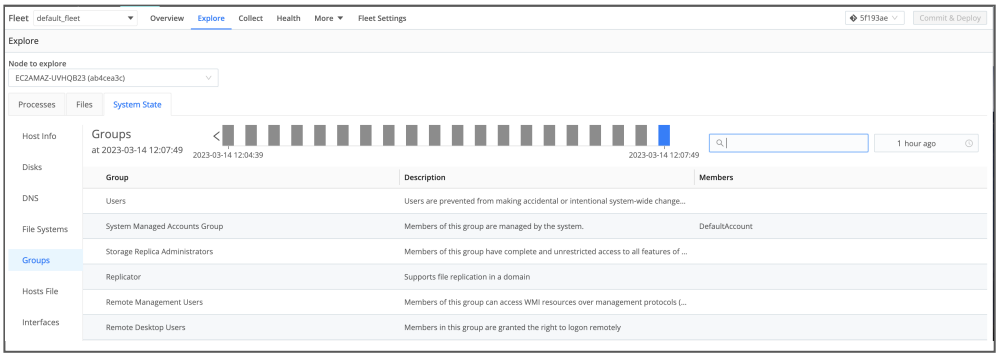
The **Firewall** tab displays a list of the host's defined firewall rules.



Firewall tab

Groups

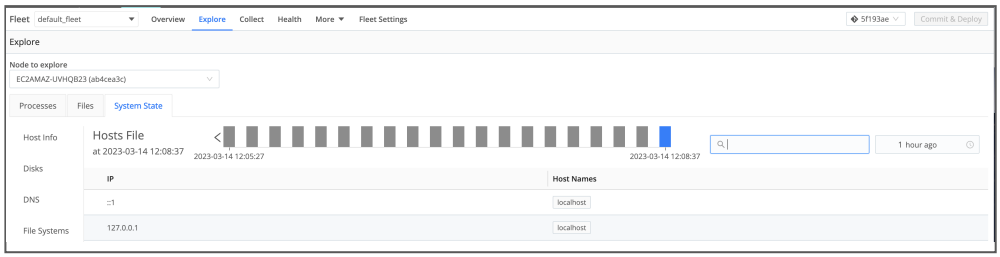
The **Groups** tab displays a list of local groups including their names, descriptions, and members on the host system.



Groups tab

Hosts File

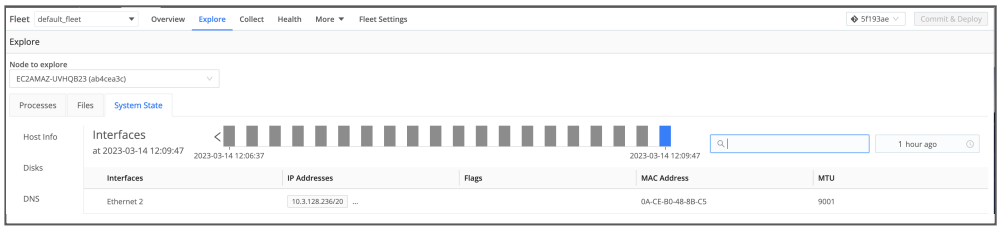
The **Hosts File** tab displays the current state on the host system.



Hosts File tab

Interfaces

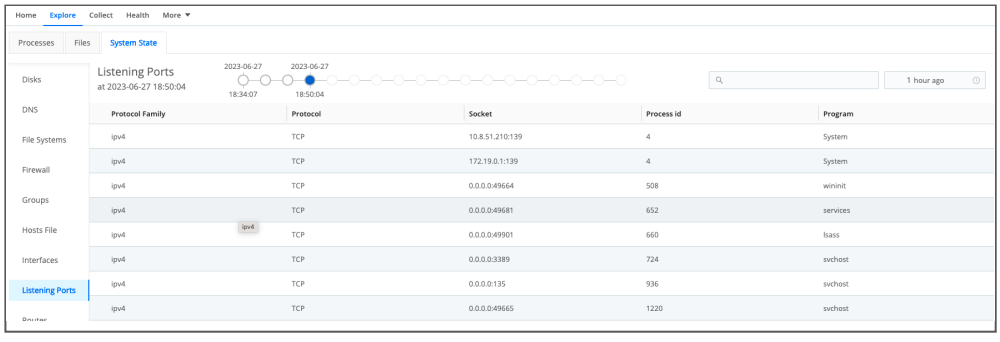
The **Interfaces** tab displays a list of each of the network interfaces on the host system.



Interfaces tab

Listening Ports

The **Listening Ports** tab displays a list of listening ports and their associated process identifier (pid).



Listening Ports tab

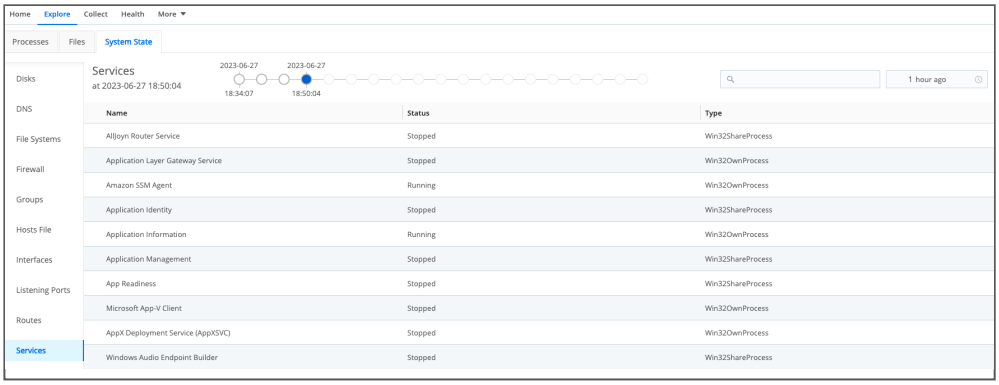
Logged-In Users

The **Logged-In Users** tab displays a list of currently logged-in users on the host.

Logged-In Users tab

Services

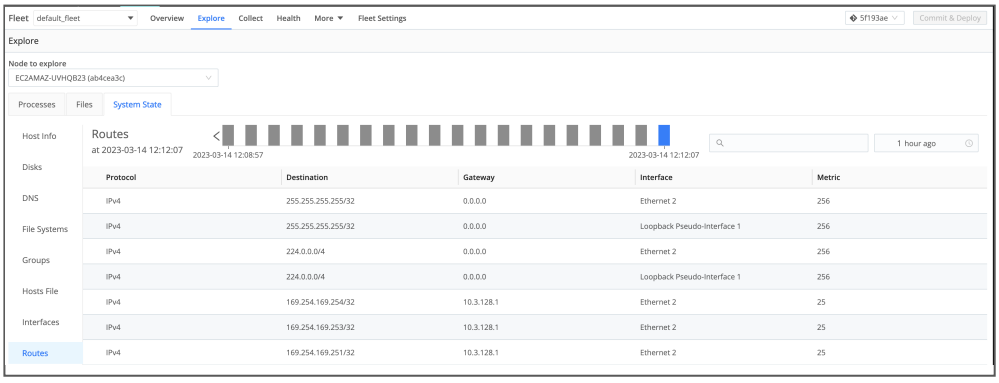
The **Services** tab displays a list of each configured service along with their running status.



Services tab

Routes

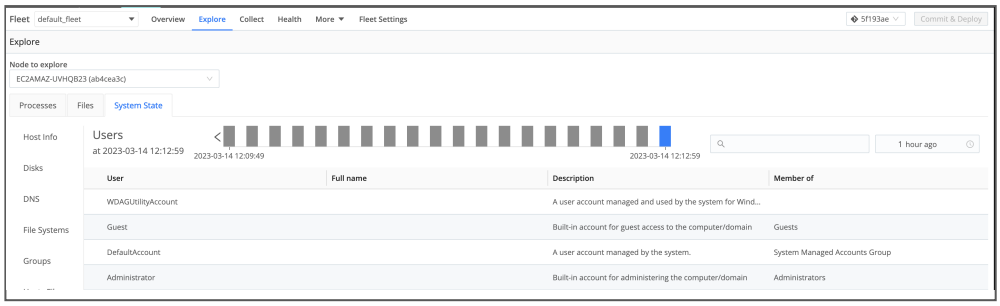
The **Routes** tab displays entries from network routes on the host system.



Hosts Routes tab

Users

The **Users** tab displays a list of local users on the host system.



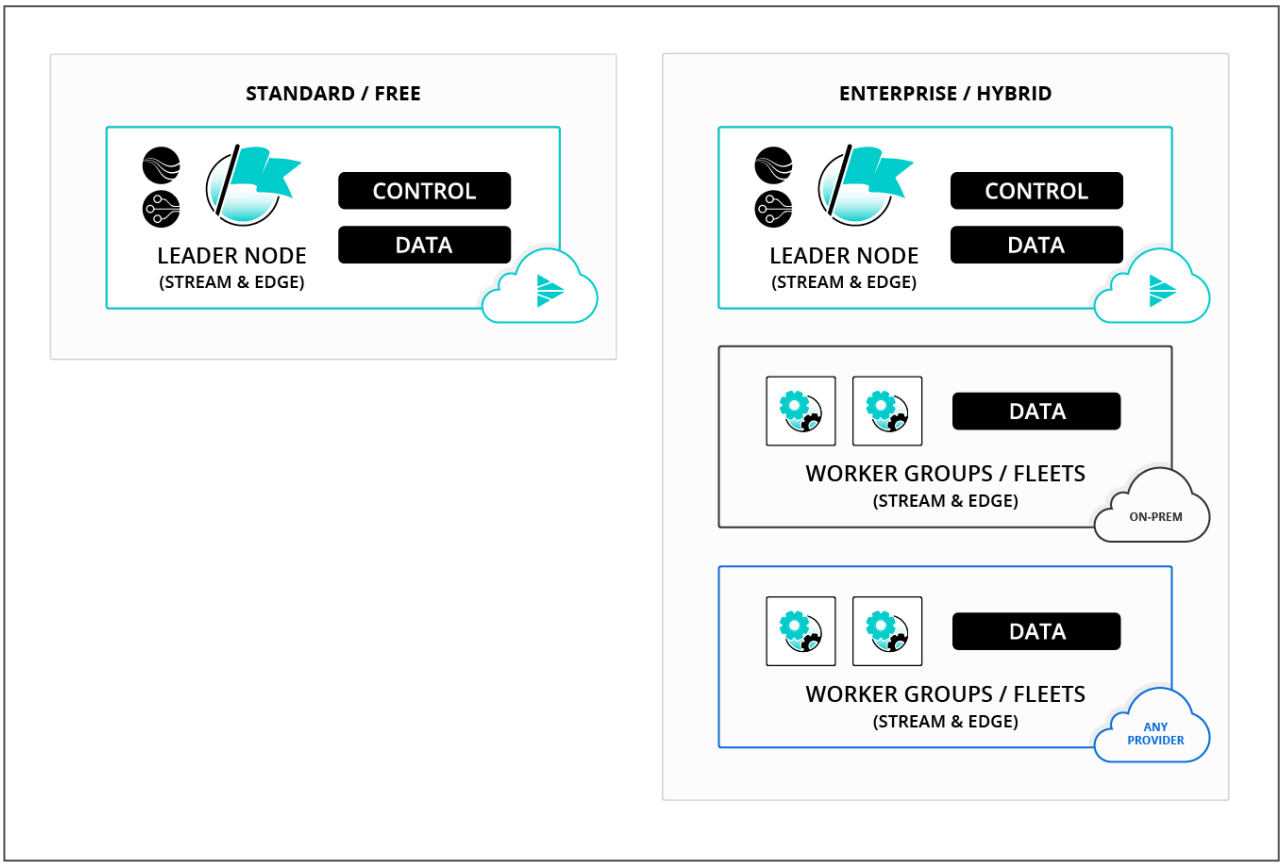
Users tab

Cribl University offers a course titled Collecting Data in Windows that provides a good overview of working with Windows. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

2. Cribl.Cloud

The fast alternative to downloading and self-hosting Cribl Edge software is to launch Cribl.Cloud. This SaaS version, which comes in both a free and a paid version, places the Leader and the Edge Node in Cribl.Cloud, where Cribl assumes responsibility for managing the infrastructure.

By upgrading to a Cribl.Cloud Enterprise plan, you can implement a hybrid deployment of any complexity. In hybrid deployments, the Leader (the control plane) resides in Cribl.Cloud, while the Workers that process the data (the data plane) can reside in any combination of Cribl-managed Workers/Edge Nodes, on-prem or private cloud instances that you manage, and your data centers.



Standard/free versus Enterprise/hybrid deployment

For an overview of additional features available on Enterprise plans, see Pricing.

Why Use Cloud Deployment?

Cribl.Cloud is designed to simplify deployment, and to provide certain advantages over using your own infrastructure, in exchange for some current restrictions (because Cribl will manage some configuration on your behalf). Cribl.Cloud offers, among others:

- Full Cribl Edge power with no responsibility to install or manage software. Cribl.Cloud is fully hosted and managed by Cribl, so you can launch a configured instance within minutes.
- Access to Cribl Search to search, explore, and analyze machine data in place.
- Automated delivery of upgrades and new features.
- Encrypted data at rest (configuration, sample files, etc.) at the disk level for Leader and Cribl-managed Worker instances.
- Free, up to 1 TB/day of data throughput (data ingress + egress) for all new accounts.
- Quick expansion of your Cribl.Cloud deployment beyond the free tier's limits by purchasing credits toward metered billing. Pay only for what you use.

If you're new to Cribl Edge, please see our Basic Concepts page and Getting Started Guide for orientation. This Cribl.Cloud documentation focuses on a Cloud deployment's differences from other deployment options – referred to as "customer-managed deployments."

Cribl.Cloud always runs in distributed mode – see Simplified Distributed Architecture for details.

Cloud Pricing

Beyond the free tier, an optional paid Cribl.Cloud account – whether Standard or Enterprise – offers direct support, plus expanded daily data throughput according to your needs. From your Cribl.Cloud Organization, select **Go Enterprise** to submit an inquiry about upgrading your free account, and Cribl will respond.

You'll pay only for what you use – the data you send to Cribl Edge, and the data sent to external destinations. However, data sent to your AWS S3 storage is always free. For details, see Pricing.

2.1. Initial Cribl.Cloud Setup

Your first step to using Cribl.Cloud is to sign up on the Cribl.Cloud portal (see Registering a Cribl.Cloud Portal below), to create your Cribl.Cloud Organization.

Your Organization will display a dedicated portal, a network and access boundary that isolates your Cribl resources from all other users. Each Cribl.Cloud account provisions a separate AWS account. Your instances of Cribl Stream, Cribl Edge, and Cribl Search are deployed inside a virtual private cloud (VPC) in this account.

The portal will initially be on a **free** Cribl.Cloud plan. Certain throughput and administration limits apply to a free account. When you need more capacity and/or options, it's easy to upgrade to a paid or Enterprise plan – just click the **Go Enterprise** button at the top of your portal.

The Cribl.Cloud Suite is listed on AWS Marketplace. When you're ready for a paid plan, you can use your Enterprise Discount Program (EDP) credits here to run Cribl products, billed through your AWS account – with no need for a separate procurement process.

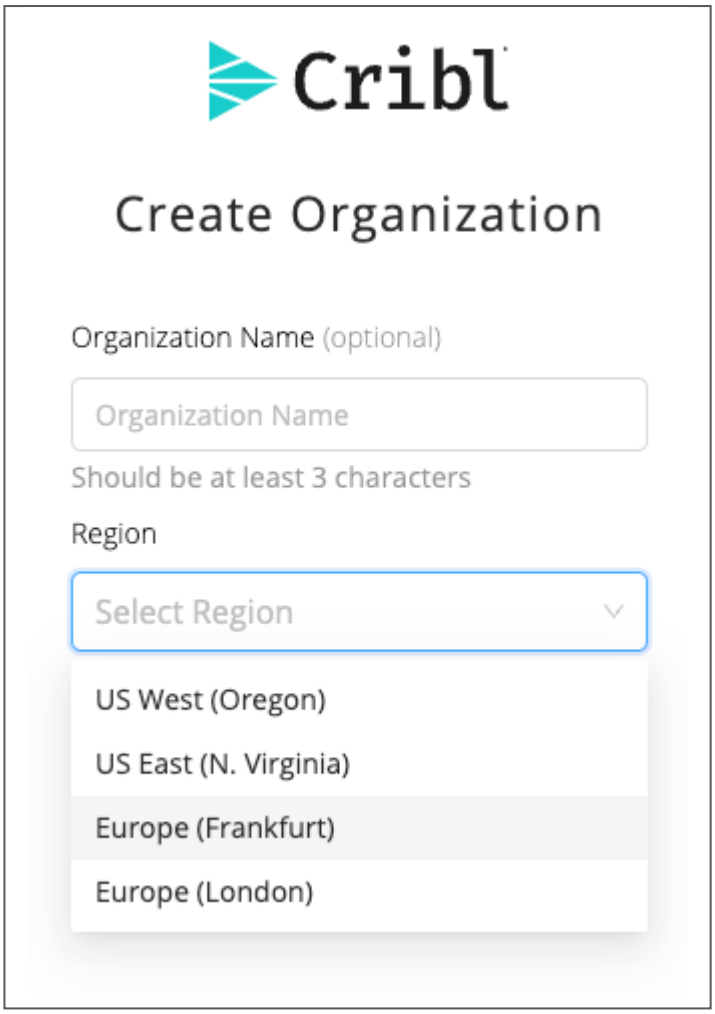
Cribl is SOC 2 (Service Organization Control 2) Type II security compliance certified.

Registering a Cribl.Cloud Portal

Ready to take the plunge? Here's how to register and manage a Cribl.Cloud instance.

First, if you haven't already signed up on Cribl.Cloud:

1. Start at: https://cribl.cloud/signup/
2. Register with your work email address.
3. Use the verification code from Cribl's email to confirm your registration.
4. On the **Create Organization** page, optionally enter an **Organization Name** (a friendly alias for the randomly generated ID that Cribl will assign to your Organization).
5. Select an AWS Region to host your Cribl.Cloud Leader and Cribl-managed Workers. Cribl currently supports the following Regions:
    - `US West (Oregon)`
    - `US East (Virginia)`
    - `Europe (Frankfurt)`
    - `Europe (London)`
6. Bookmark your Cribl.Cloud portal page, for all that follows.



Create Organization page – selecting a host Region

Note that each user can register only one active Organization. For details, see notes on Member Permissions.

Troubleshooting Resources

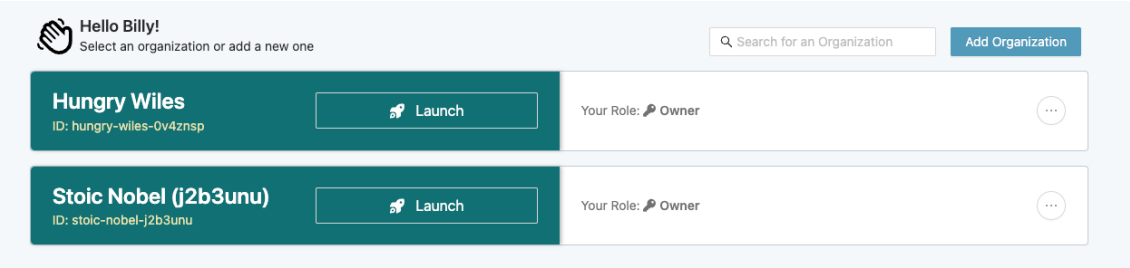
Cribl University's Troubleshooting Criblet on How to Log Into Cribl.Cloud walks you through the whole registration and login flow. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Troubleshooting Criblets and Advanced Troubleshooting short courses.

2.2. Cribl.Cloud Portal

The Cribl.Cloud portal is the place where you get an overview of and manage your organization, configure network settings, and control user access and billing information.

Select Organization Page

When you own or are a member of multiple Cribl.Cloud Organizations, the Organization selection page – displayed after you first sign in, or sign back in after a logout – enables you to choose which Organization you want to work with. You can later switch your organization via the Account menu.



Select Organization interstitial page

Click any tile's `...` button to reveal an options menu. Here you can check who is the owner of the Organization. You can also click **Leave Organization** if you want to remove yourself as a member of another owner's Organization. This option requires confirmation – proceed only if you're sure! (You won't see this button on Organizations that you own.)

Managing Cribl.Cloud

Once you've registered on the portal, here's how to access Cribl.Cloud:

1. Sign in to your Cribl.Cloud portal page.
2. Select the Organization to work with.
3. From the portal page, select **Manage Stream**, **Manage Edge**, or **Explore [Search]**.
4. The selected application's UI will open in a new tab or window – ready to goat!

Note the **Cribl.Cloud** link at the Cribl.Cloud home page's upper left, under the **Welcome!** message. You can click this link to reopen the Cribl.Cloud portal page and all its resources.

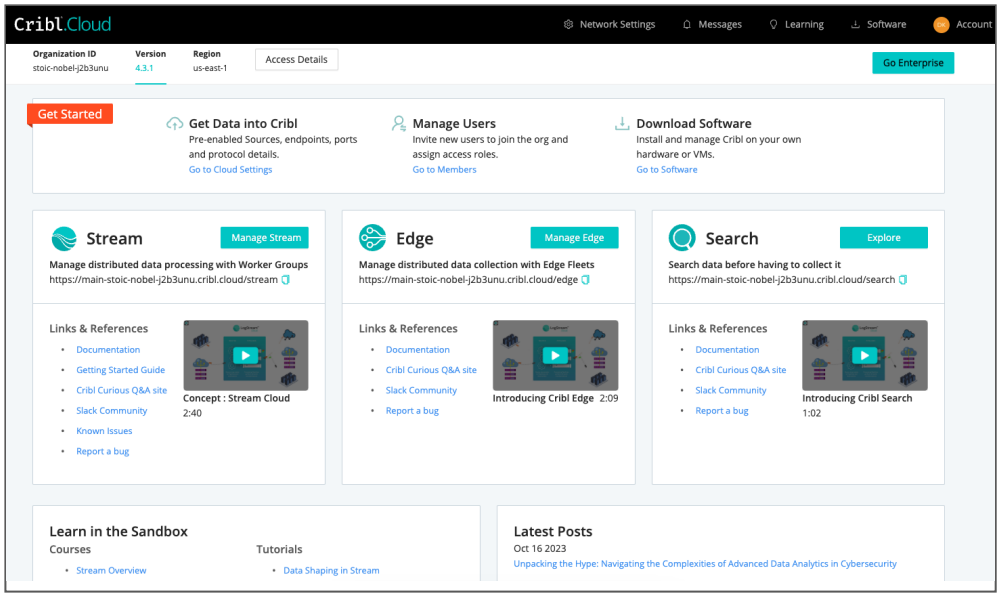
Exploring the Cribl.Cloud Portal

Now that you're here – explore the furniture. The Cribl.Cloud portal's top navigation allows you to navigate among the following pages/links:

- Portal (**Cribl.Cloud** logo)
- Network Settings
- Messages
- Learning
- Software
- Account (including Organization details)

Portal Page

When you log into the Cribl.Cloud portal, you'll land here. The main events here are the **Manage Stream**, **Manage Edge**, and **Explore [Search]** buttons. Click these to launch (respectively) Cribl Stream, Cribl Edge, or Cribl Search in a new tab.



Cribl.Cloud portal

However, the surrounding page offers lots more useful information:

- On the page body, you'll find links to multiple Cribl resources – documentation, support (Community Slack and bug reporting), free Sandbox training, and blog posts.
- In the Overview strip just below the top black menu, you'll find detailed configuration information about your Cribl.Cloud Organization.
- By clicking the top nav's **Network Settings** link, you can check and manage connectivity details – data Sources, access control, and trust relationships – for your Cribl-managed Cribl.Cloud Workers.

Overview and Access Details

From left to right, this upper strip displays the following config details:

Organization ID: Domain at which you access the associated Cribl.Cloud Organization.

Version: The version of Cribl Stream/Edge applications deployed to your Organization and its Cribl-managed Workers.

Region: The AWS Region where you're running Cribl applications. (Cribl.Cloud currently supports the following Regions: **us-west-2**, **us-east-1**, **eu-central-1**, and **eu-west-2**.)

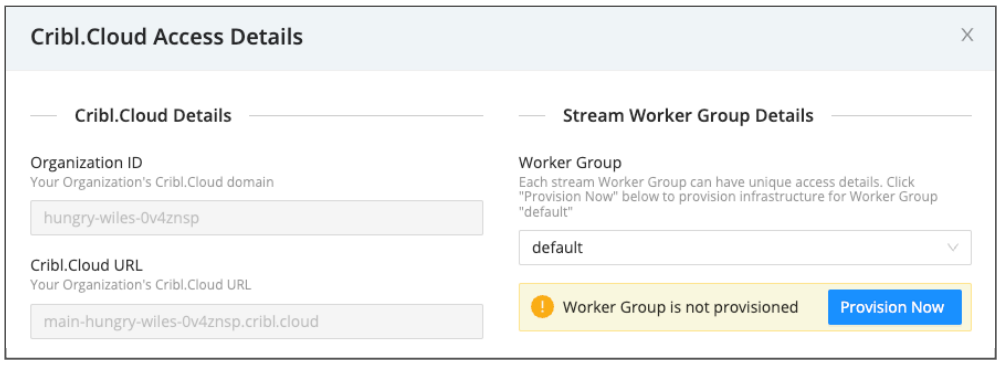
Access Details: See your Organization's details.

The left column repeats the Organization, region, and version information, with one addition:

Cribl.Cloud URL: Static address associated with the load balancer that is in front of the Leader. Hybrid Edge Nodes will connect to this address on port 4200, while the Leader UI is served from this address on port 443.

The right column provides a consolidated, read-only display of the following **Stream Worker Group Details**. Some of these options are configured on different tabs, as noted below.

- **Worker Group**: Use this drop-down to select any Group of Cribl-managed Workers that you've configured (including `default`). The remaining fields on the right will display details specific to that Group.
- **Provision Now**: This button will replace all the fields listed below when a Group is dormant. Click the button when you're ready to provision infrastructure for the Group. After a lag, the Group will be ready to process data, and this modal's remaining fields will populate.



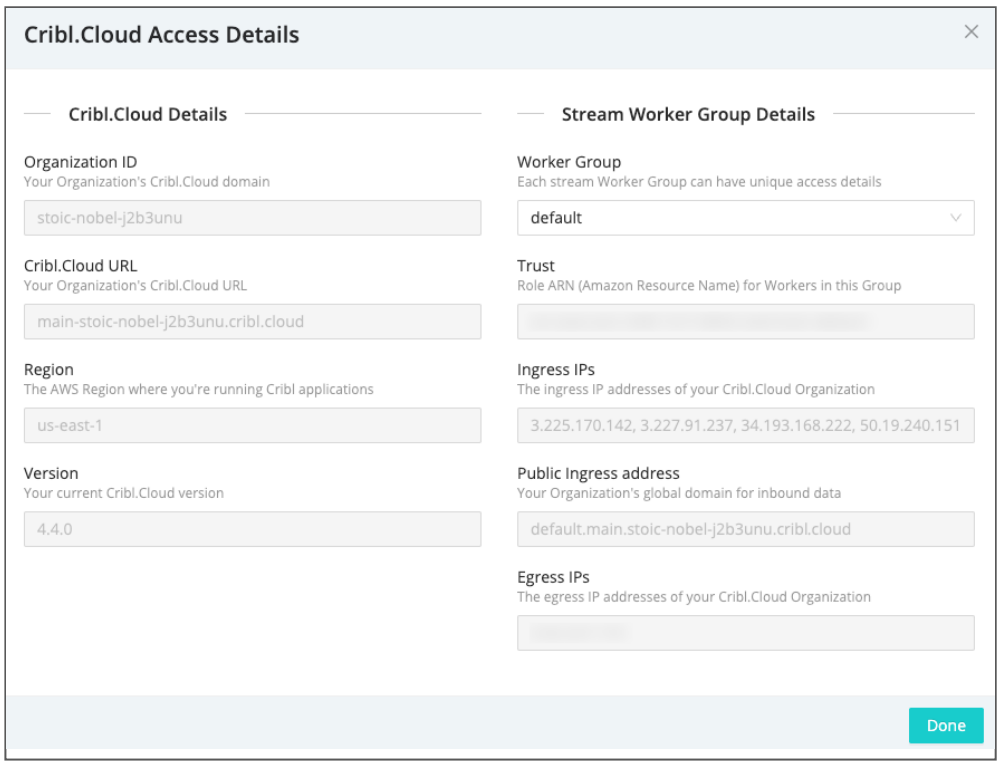
Access Details modal for an unprovisioned Group

- **Trust**: Role ARN for Workers in this Group. You configure these ARNs on the Trust tab.
- **Ingress IPs**: The IPv4 addresses of Edge Nodes' load balancers, also used when receiving data from Push Sources. These addresses will remain constant, so you can build firewall rules around them.

- **Public Ingress address**: Each Group's domain for inbound data. This address prepends the Group name to the Organization's global domain name. It does not append ports per data type – you can obtain these from the Data Sources tab.
- **Egress IPs**: Your Cribl.Cloud Organization's current public IP addresses. These addresses are Group-specific and also dynamic: Cribl will occasionally update them when we need to rescale core infrastructure. The addresses are used for both outbound connections from the Workers and Pull Sources.

Egress IPs are not static.

If you need a static egress IP, contact Cribl Support.



Access Details modal for a provisioned Group

Configuring Stream Groups (beyond the `default` Group) requires an Enterprise plan. For details about creating and provisioning Groups, see Cribl.Cloud Worker Groups.

The **Access Details** modal's left side displays Organization-wide access details, including the **Cribl.Cloud URL** of your Org's Leader/control pane. You'd use this URL for certain API calls and certain Collection operations coordinated by the Leader. Use the right-side details to configure data flow through individual Groups.

Network Settings

Clicking the top nav's **Network Settings** link opens a page with connectivity details, spread across three upper tabs: **Data Sources**, **Trust**, and **ACL**.

Data Sources

The **Data Sources** tab lists ports, protocols, and data ingestion inputs that are open and available to use, including pre-enabled Sources. Use the **Group** drop-down to filter these details per Group of Cribl-managed Workers in the Stream app. Return to this tab to copy Ingest Addresses (endpoints) as needed. For details, see Available Ports and TLS Configurations.

For each existing Source listed here, Cribl recommends using the preconfigured endpoint and port to send data into Cribl Stream.

Trust

The **Trust** tab provides **Worker ARN**s (Amazon Resource Names) that you can copy and paste to attach a Trust Relationship to an AWS account's IAM role. Use the **Group** drop-down to display the ARN for any Group of Cribl-managed Stream Workers.

Attaching a Trust Relationship enables the `AssumeRole` action, providing cross-account access. For usage details, see the AWS Cross-Account Data Collection topic's **Account B Configuration** section.

This option applies only to your Cribl-managed Workers. You cannot use this technique to enable access to hybrid Workers on customer-managed Cribl Edge instances.

ACL

This **Access Control List** defines Rules (IPv4 CIDR ranges) to restrict data sent to your data sources. The Rules you define here are global to all your Cribl-managed Groups of Stream Workers.

The default `0.0.0.0/0` rule (modifiable) imposes no limits. Click `+` to add more rules, or click `X` to remove rules. End a rule with `/32` to specify a single IP address, or with `/24` to enable a whole CIDR block from `x.x.x.0` to `x.x.x.255`.

Click **Save** after adding, modifying, or removing rules. Each change takes up to 5 minutes to propagate. Cribl.Cloud will display an `ACL update in progress...` banner, notifying you that rules edits are temporarily disabled to prevent conflicts. A successful update proceeds silently – you will not see a confirmation message.

The **ACL** options apply only to your Cribl-managed Workers. You cannot use this technique to set

> access rules on hybrid Workers running in customer-managed Cribl Edge instances.

Messages

Clicking the top nav's **Messages** link opens the **Message Center** right drawer. Here, you will find Cribl.Cloud status and update notifications from Cribl, with **Unread** messages above the **Read** group.

Learning

Clicking the top nav's **Learning** link opens the **Learning** page, which provides links to everything you need to learn about Cribl Edge in order to goat forth and do great things:

- Sandboxes (free, interactive tutorials on fully hosted integrations).
- Documentation.
- Product and plans overview (pricing comparison).
- Cribl events (including future and archived Webinars).
- Concept/demo videos.

Software

If you want to try an on-prem installation, this page offers download links for Cribl Edge, Cribl Edge, and AppScope software. You can download either binary installation files or Docker containers (hosting Ubuntu 20.04), to install and manage on your own hardware or virtual machines.

Account

Manage your Cribl.Cloud account with these submenu items. Hover over **Account** to reveal more options:

- **Profile** allows you to update your personal information.
- **Organization** provides details about the current organization (for an Organization's owner only).
- **Organization Selection** submenu (fly-out) works like the Select Organization page – select a link to traverse to other Organizations.
- **Log Out** signs you out of the account you're on, and takes you back to the login screen.

Organization

Displayed only to an Organization's owner, this page offers information about your Organization's details, Members, and (where applicable) billing and SSO, organized into tabs along the top of the page.

Access to the **Billing** tab includes the **Plan & Invoices** and **Usage** left tabs.

Details Tab

You can add these optional details to make your Cribl.Cloud deployment more recognizable than its randomly generated **Organization ID**:

Field name	Description
Alias	A "friendly" name for your Organization. Upon signing in, members will see this alias above the Organization ID on the Select Organization page.
Description	Add further details about your Organization.
Opt in to beta features	If displayed, this toggle enables access to new options that Cribl has not yet made generally available. As with all beta features, expect some instability in exchange for advancing to the cutting edge of your Cribl.Cloud.

Click **Save** to immediately apply your changes.

Members

This tab provides access to inviting and managing other users.

API Management

This tab provides access to existing API credentials and the ability to create new ones.

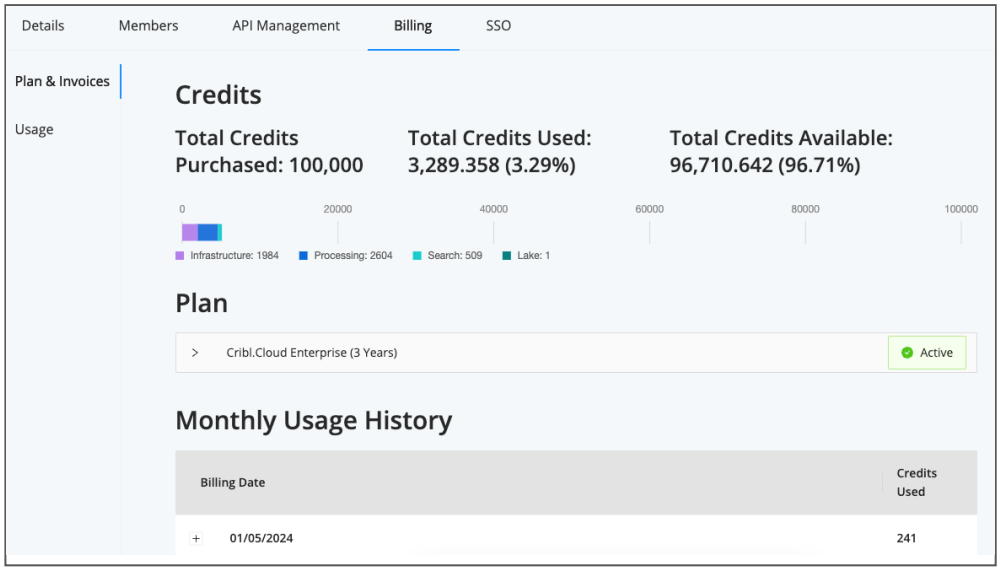
Billing

This tab is displayed only to owners of an Organization on a paid license plan. It provides Plan and Usage left tabs:

Plan & Invoices

The **Plan & Invoices** left tab displays a mercury bar of purchased, used, and available **Credits** on your account. Color-coding breaks down usage by infrastructure versus processing (data throughput).

Below that is an expandable **Plan** details section. Expandable **Monthly Usage History** rows offer details about your credits usage in the current and prior months. Here, you can break out usage on Cribl Search, on hybrid Workers (ingest billing only), and on Groups of Cribl-managed Workers (with ingest versus infrastructure breakouts per Group).



Billing > Plan & Invoices tab

Credits carry over across billing periods, as long as you renew your Cribl.Cloud plan.

Usage

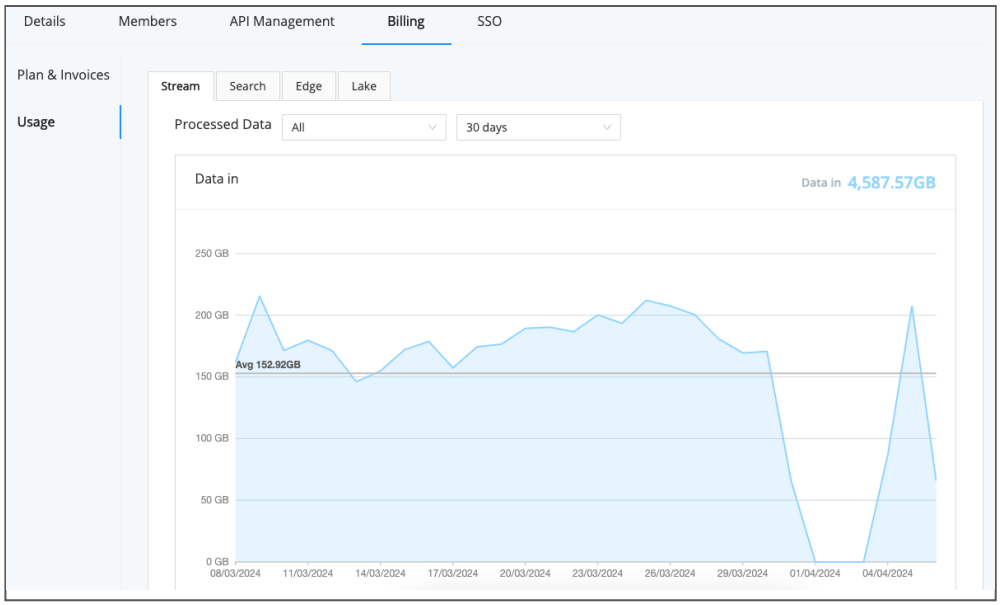
The **Usage** left tab provides nested tabs for **Stream**, **Search**, **Edge**, and **Lake**. (If you're keeping tabs, this is a third level of selectable tabs.)

Select a product to view its graphs, and adjust the duration of time for which data was processed over a selectable trailing period of 7 days to 1 year.

Stream	Search	Lake	Edge
The **Stream** tab graphs credits usage for **Data in**, **Data out**, and **Infrastructure**. Using the **Processed Data** drop-down, you can filter the aggregate display down to individual Cribl-managed Groups, or to all hybrid Groups.	The **Search** tab graphs billed **Search Compute** in CPU hours and **Total Compute Costs**.	The **Lake** tab graphs **Total Usage** and **Data Usage**.	The **Edge** tab graphs **Data Ingest** and **Data Egress**.

The trend line shows daily averages, and you can hover over data points to pop out details.

At the right side of each graph is a total for the selected period.



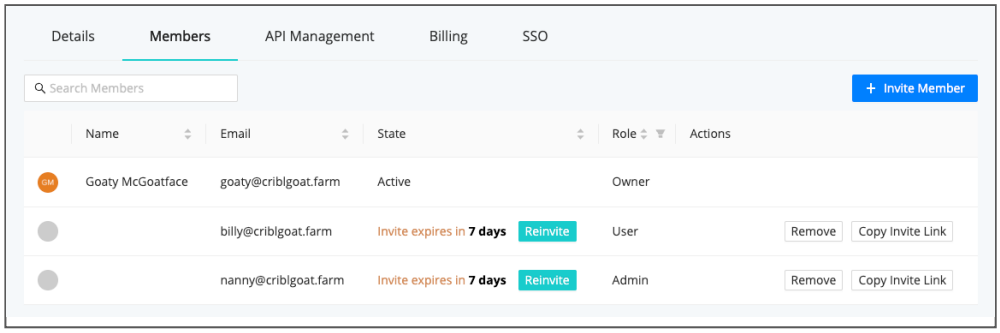
Billing > Usage tab

SSO Tab

This tab appears on an Enterprise plan, enabling you to configure federated authentication to your Cribl.Cloud Organization from an OIDC or SAML identity provider. For details, see Cribl.Cloud SSO Setup.

2.3. Managing Cribl.Cloud

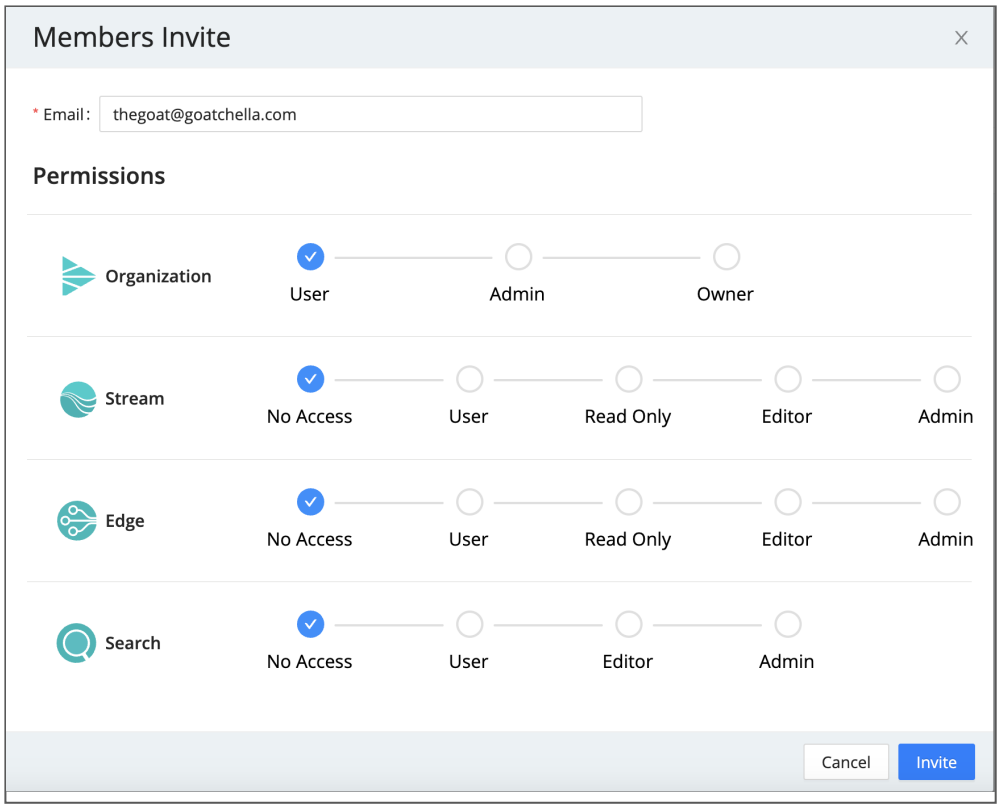
From the **Organization** > **Members** tab, an Organization's owner can invite new users to join the Organization, assign access permissions to new and existing members, remove pending invites, and remove existing members.



Organization > Members tab: Managing Invites and Members

Inviting Members

Click **Invite Member** to open the modal shown below. Enter the **Email address** of the new user you want to invite, assign them appropriate **permission** on your Organization, and then click **Invite** to send the invitation.



Invite User modal

Responding to Invites

At the address you entered, the new member receives an email with an **Accept Invitation** link to either sign into their existing Cribl.Cloud account, or else sign up to create an account and its credentials.

After signing in, they'll have access to your Organization and Cribl Edge instance at the permission level you've specified.

Organization Member Permissions

Newly configured Members start out with the permissions they were granted when they were invited at the Organization level, as well as the **No Access** permission on each product. An Organization's Owner can assign the **User**, **Admin**, and **Owner** permissions at the Organization level. Assigning the **User** permission requires an Enterprise license.

You assign permissions per individual user when you invite them to your Organization, or after they join it. Cribl.Cloud does not currently support globally predefining or assigning group permissions, as with on-prem Cribl Edge.

For permissions that you can assign at the lower product, Fleet, and resource levels (all of those only with an Enterprise plan), see our Members and Permissions topic, which also covers on-prem counterparts to the Cribl.Cloud Organization-level permissions listed here.

Permission	User	Admin	Owner
Log into the system			
Update own member profile			
View Stream Worker groups you have access to			
`Read Only` Access to all Products			
`Admin` Access to all Products			
View and execute Leader commits			
View and modify Global Settings			
Manage ACLs (Access Control lists)			
View and update SSO settings			
View Data Sources and Trust Policies			

Permission	User	Admin	Owner
Manage API Credentials			
View and manage Cribl Suite Global Settings			
View and manage (provision, update, and delete) Stream Worker Groups			
View and execute Leader commits			
View the Organization's Billing dashboards			
View Organization Details			
Update Organization Details			
Delete an Organization			
View Organization members			
Manage (invite, update, delete) Organization members			

Each user can have Owner permissions for multiple organizations, and each organization can have multiple users as Owners. However, each user – as defined by their email address – can register only one active Organization, and only if they are not already the Owner of a different Organization.

Cribl.Cloud has retired the access-management model of Local Users and Roles/Policies that Cribl applications used prior to v.4.2. Create and configure only **Members** on Cribl applications – references elsewhere in Cribl docs to Local Users do not apply to Cribl.Cloud.

Managing Invites

While an invite is pending, the **Organization** > **Members** tab offers you these options to deal with commonly encountered issues:

- **Reinvite**: If your invited member didn't receive your invitation email, you can click this button to resend it.
- **Copy Link**: If emails aren't getting through at all, click this button to copy and share a URL that will take the invitee directly to the signup page. This target page encapsulates the same identity, Organization, and permission you specified in the original email invite.
- **Remove**: This is for scenarios where you need to revoke a pending invite. (You sent someone a duplicate invite, your invitee is spending too much time in space to be a productive collaborator, etc.) After clicking this button, you'll see a confirmation dialog.

After seven days, if an invite has been neither accepted nor revoked, it expires. In this case, it is removed from the **Members** tab.

Managing Invites

Managing Members

Once a user has accepted an invite, the **Organization** > **Members** tab offers you these options to modify their membership in your Organization:

- **Edit**: Switch this member to a different Permission. (The **Edit** option is displayed only if you have an Enterprise plan.)
- **Remove**: Remove this member from your Organization. After clicking this button, you'll see a confirmation dialog. (Proceeding will not affect this user's access to any other Cribl.Cloud Organizations they might own or be members of.)

2.4. Cribl.Cloud Vs. Self-Hosted

A Cribl.Cloud deployment can differ from an on-prem/customer-managed Cribl Edge deployment. The following documentation lists the main ways in which the Free tier of Cribl.Cloud diverges from a customer-managed deployment. Keep in mind all these differences as you navigate Cribl Edge's current UI, in-app help (including tooltips), and documentation.

Simplified Administration

Cribl.Cloud has been designed with options to accommodate everyone – from first-time evaluators, to Enterprise customers managing a worldwide network of private-cloud, public-cloud, and/or data-center deployments.

Cribl.Cloud's free offering is designed to help you launch Cribl Edge – and to start processing data – as quickly and easily as possible. Cribl manages many features on your behalf, allowing for a streamlined Settings left nav.

Below are the key options streamlined out of the free Cloud offering. Bear in mind that upgrading to an Enterprise plan will make many of these options configurable:

Simplified Distributed Architecture

Cribl.Cloud is preconfigured as a distributed deployment for Cribl Stream or Cribl Edge. With a Free or Standard plan, allows only a single Fleet.

Compared to self-hosted Cribl Edge, the **Settings** > **Worker Processes** and **Settings** > **Distributed Settings** links are omitted.

With an Enterprise plan, Cribl always provides at least two Workers, and will scale up further Workers as needed to meet your peak load. With an Enterprise plan, you also have the option to configure additional hybrid Edge Nodes and Fleets.

Git Preconfigured

Without an Enterprise plan, the **Settings** > **Global Settings** > **System** > **Git Settings** section is omitted. A local `git` client is preconfigured in your Cribl.Cloud portal. On Cribl.Cloud's top nav, use the **Global Config** link (branched icon) to commit/push changes to `git`. Select **Deploy** to deploy your committed changes. Cribl.Cloud does not support Git remote repos.

Automatic Restarts and Upgrades

Without an Enterprise plan, the **Settings** > **Controls** and **Settings** > **Upgrade** links are omitted. Cribl handles restarts and version upgrades automatically on your behalf.

Simplified Access Management and Security

In Cribl.Cloud, you can manage access control for your Organization by clicking **Account** > **Organization** and selecting the **Members** tab. The options on this tab will vary depending on your plan.

If you have a Cribl.Cloud Enterprise plan, you can use the Key Management Service (KMS), which maintains the keys Cribl Edge uses to encrypt secrets on Fleets and Edge Nodes. Go to **Settings** > **Security** > **KMS** to configure KMS.

If you add an Enterprise Plan, cloud and hybrid Leaders support Local and Google SSO authentication, along with OpenID Connect (OIDC) and SAML federated authentication. Cribl.Cloud does not currently support LDAP.

Permission- and Role-based access control (RBAC) is simplified in Cribl.Cloud. For details, see Member Permissions.

Transparent Licensing

The top nav's **Settings** > **Global Settings** > **Licensing** link is omitted. Your license is managed by your parent Cribl.Cloud portal, where you can check credits and usage history on the Billing tab.

Other Simplified Settings

Cribl is gradually narrowing the limitations listed in this section, as Cribl.Cloud gains feature parity with on-prem deployments.

Available only on hybrid, customer-managed Edge Nodes:

- Script Collector
- Staging directory support in File-based Destinations
- Tee Function
- File System Collector
- Filesystem Destination

Available only on self-hosted deployments:

- Top nav's **Settings** > **Global Settings** > **Scripts** (Cribl.Cloud currently does not support configuring or running shell scripts on hybrid or Cribl-managed Edge Nodes.)

Available on Cribl-managed Edge Nodes, but unavailable on Cribl-managed Workers:

- System State Source
- AppScope Source's **Filter Settings**

Persistent Queues can be configured on both hybrid and Cribl-managed Edge Nodes, with an Enterprise plan. On hybrid Edge Nodes, you can freely define the **Max queue size**, based on the disk space you provision. On Cribl-managed Edge Nodes, each Source or Destination's queue is allocated a maximum of 1 GB disk space per Worker Process. (Given this automatic configuration, Cribl-managed Sources and Destinations expose only limited PQ controls.)

Support Options

At **Settings** > **Diagnostics**, you can generate diagnostic bundles and send them directly to Cribl Support. Currently, you cannot download diags. For all support options, see Working with Cribl Support.

Available Ports and TLS Configurations

To get data into Cribl.Cloud, your Cribl.Cloud portal provides several Sources and ports already enabled for you, plus 11 additional TCP ports (`20000-20010`) that you can use to add and configure more Cribl Edge Sources.

TLS Details

TLS encryption is also pre-enabled for you on several Sources, also indicated on the Cribl.Cloud portal's **Data Sources** tab.

Cribl HTTP and Cribl TCP Sources/Destinations

Use the Cribl HTTP Destination and Source, and/or the Cribl TCP Destination and Source, to relay data between Edge Nodes connected to the same Leader. This traffic does not count against your ingestion quota, so this routing prevents double-billing. (For related details, see Exemptions from License Quotas.)

Simplified Source, Collector, and Destination Configuration

Several commonly used Sources are preconfigured for you within Cribl.Cloud's UI, and are ready to use.

The Exec Source is unavailable on Cribl-managed Workers, but is available on hybrid Workers.

The Cribl Internal Source's CriblLogs option is unavailable in Cribl-managed Stream instances, but it is available in Cribl Edge, and in hybrid Workers' Stream instances. The Cribl Internal > CriblMetrics option is available in all of the above combinations.

2.5. Enterprise Cloud

With a Cribl.Cloud Enterprise plan, you have the same options and flexibility that an Enterprise license provides for a customer-managed (on-prem) distributed deployment – and more:

- Configuring and managing multiple Worker Groups and Fleets.
- Notifications within Cribl apps, to PagerDuty, and/or to other services via webhook.
- Fine-grained, role-based control of Member Permissions on your Cribl.Cloud Organization and on individual product resources.
- Single sign-on/SSO authentication from external identity providers.
- The hybrid deployment option, described just below.

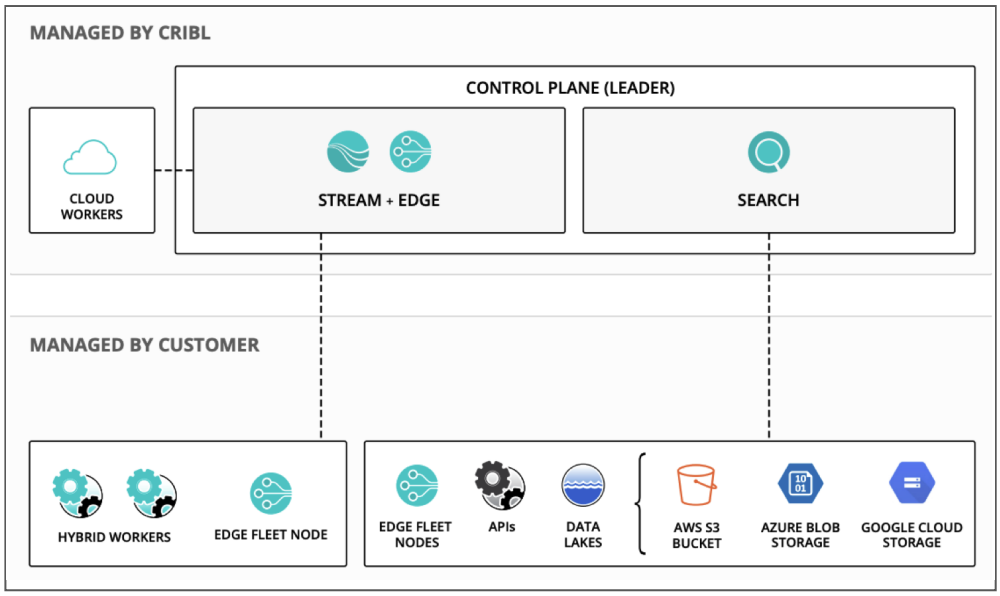
With an Enterprise Plan, the Leader resides in Cribl.Cloud, and controls a flexible mix of Cribl-managed and/or customer-managed Fleets. Cribl manages the Leader's high availability on your behalf.

For other Enterprise features – and for comparisons between Cribl.Cloud plans and on-prem licenses – see Cribl's Pricing page.

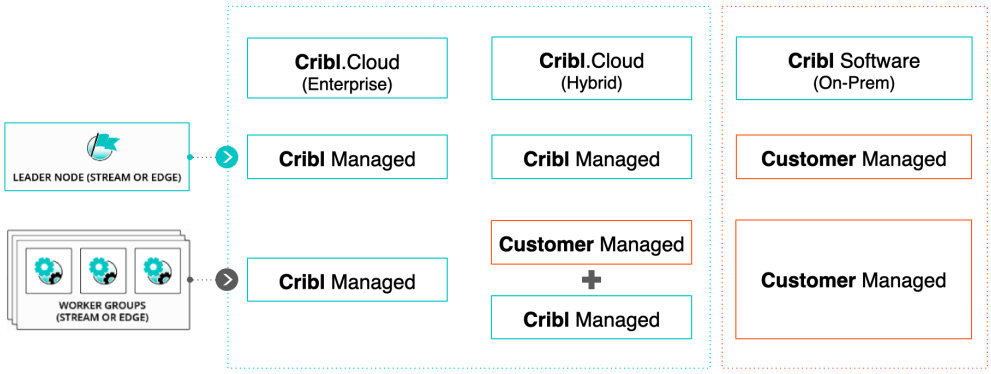
Hybrid Deployment

The diagrams below show the comparative flexibility of a hybrid Cribl.Cloud deployment. The Leader (control plane) resides in Cribl.Cloud, while the Workers that process the data can be in any combination of the following environments:

- In Cribl.Cloud, managed by Cribl.
- In public or private cloud instances that you manage.
- On-premises in your data centers.



Enterprise hybrid deployment, with control plane and Cribl-managed Workers in Cribl.Cloud



Enterprise hybrid deployment, with only control plane in Cribl.Cloud

As the footprint of your operations grows or changes, this flexibility makes it easy to reconfigure Cribl Edge in tandem. You can rapidly expand Cribl Edge observability into new cloud regions – and replace monitored hardware data centers with cloud instances – all while maintaining one centralized point of control.

You can also add Workers or Edge Nodes, and reassign them to different Fleets, by easily auto-generating Stream or Edge command-line scripts within Cribl Edge's UI.

Hybrid Requirements

A hybrid deployment imposes these configuration requirements:

- Hybrid Workers (meaning, Workers that you deploy on-prem, or in cloud instances that you yourself manage) must be assigned to a different Fleet than the Cribl-managed `default` Group – which can contain its own Edge Nodes.

- All Edge Nodes' hosts must allow outbound communication to the Cribl.Cloud Leader's port 4200 at `https://main-<Organization-name>.cribl.cloud:4200`, to enable configuration and workload management by the Leader.
- On all Edge Nodes' hosts, firewalls must allow outbound communication on port 443 to the Leader and to `https://cdn.cribl.io`. This port is also used to bootstrap hybrid Workers from the Leader.
- All Edge Nodes require connectivity to `https://cdn.cribl.io/telemetry/`. For details on testing this connectivity, on the metadata transmitted to Cribl, and on how we use that data, see Telemetry Data.
- If this traffic must go through a proxy, see System Proxy Configuration for configuration details.
- To verify your Leader's Region and public URL, open the Access Details modal.

Note that you are responsible for data encryption and other security measures on Edge Node instances that you manage.

Adding (Bootstrapping) Workers

To add Workers to your hybrid Cribl.Cloud deployment, Cribl recommends that you use the script outlined in Bootstrap Workers from Leader. Hosts for the new Workers must open the same ports (4200 and 443) listed in Hybrid Requirements.

You have three options for generating the script, outlined in these subsections of the Bootstrap topic linked above:

- Auto-generate it from the Leader's UI.
- Make a `GET` API request to the Leader.
- curl the same API request.

In Cribl Edge, you access all these bootstrap options via the Manage Edge Nodes page's **Add/Update Edge Node** control.

Hybrid Cribl HTTP/TCP Configuration

If you use the Cribl HTTP Destination and Source pair, or the Cribl TCP Destination and Source pair, to relay data between Edge Nodes connected to the same Leader, configuring hybrid Workers demands particular care.

The Edge Nodes that host each pair's Destination and Source must specify exactly the same Leader Address. Otherwise, token verification will fail – breaking the connection, and preventing data flow. In hybrid Cribl.Cloud deployments, the Leader's Address format is `main-<your-Org-ID>.cribl.cloud`. When configuring a hybrid Worker, use that format in the **Address** field.

To configure hybrid Workers:

1. Log directly into their UI, then select **Settings** > **Global Settings** > **Distributed Settings**. Make sure the **Mode** is set to **Managed Worker** or **Managed Edge** (which might require a restart).
2. Then select the **Leader Settings** left tab, and ensure a consistent entry in the **Address** field.

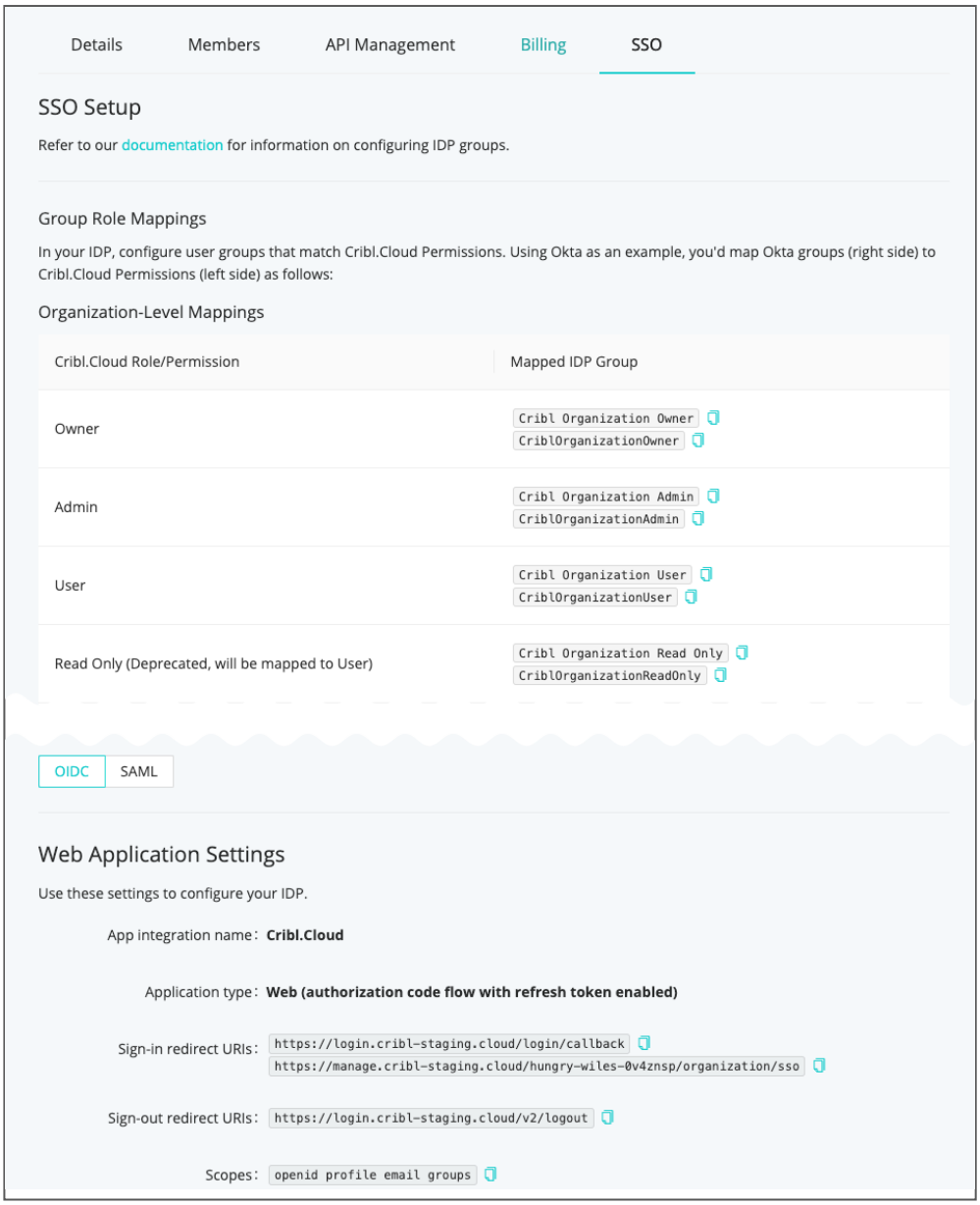
2.6. Cloud SSO Setup

2.6.1. Cribl.Cloud SSO Setup

The pages in this section outline how, with a Cribl.Cloud Enterprise plan, you can set up a Single Sign-On (SSO) integration between your identity provider and your Cribl.Cloud portal. The following pages cover both OIDC and SAML authentication options:

- Common SSO Setup Steps
- OIDC/Okta Setup Example
- SAML/Okta Setup Examples
- SAML/Microsoft Entra ID Setup Examples
- Final SSO Steps & Troubleshooting

SSO integration requires you to perform certain configuration steps in your identity provider (IDP), and to then submit corresponding information to Cribl. As of Cribl Edge 4.0, you can submit these details directly on your Cribl.Cloud portal's Organization page.



Organization page SSO tab

This section covers both sides of the process. For additional details specifically about integrating Cribl Edge with Okta, see SSO/Okta Configuration.

The general steps to set up a Single Sign-On (SSO) integration between your identity provider and your Cribl.Cloud portal are:

1. Invite at least one SSO admin to your Cribl.Cloud Organization from a fallback, separate email domain.
2. In your identity provider (IDP), configure user groups that map to Cribl.Cloud's four predefined Roles.
3. In your IDP, create an OIDC or SAML application.

If creating an OIDC application, you must use backchannel authentication. Cribl.Cloud does not support front-channel authentication via OIDC.

4. Submit your app's configuration details to Cribl on your Cribl.Cloud portal's **Organization** page > **SSO** tab. (This will complete your SSO setup on the Cribl side.)
5. In your IDP, assign groups to your users, matching the Role that each group of users should have in Cribl.Cloud.
6. In your IDP, assign the OIDC/SAML app to the Organization's owner, and to other Cribl.Cloud users.

2.6.2. Common SSO Setup Steps

This page lists initial preparation steps that are the same for all Single Sign-On (SSO) configurations, whether you're using OIDC or SAML.

Set Up Fallback Access

In your Cribl.Cloud Organization, ensure that at least one Owner creates a local account, using an email domain that's separate from the corporate domain on which you're configuring SSO. This will ensure backup access if SSO configuration breaks.

Avoiding Being Forced into SSO

By design, SSO does Home Realm Discovery (HRD), meaning that when you enter your email to log in, SSO checks the domain specified in the email address; and, if the domain matches a connected SSO provider, SSO sends you to that provider to log in.

Normally, this is a good thing. One exception is when you want to sign up for additional Cribl.Cloud orgs without being forced through the SSO for an existing Cribl.Cloud org. In that case, try the following workaround:

1. Edit the login URL to delete the word `identifier`.
2. Use the edited URL to log in; instead of forcing you through SSO, Cribl.Cloud will ask for a username and password.

For example:

- Original URL:
  `https://login.cribl.cloud/u/login/identifier?state=<long_string_of_characters>`
- Edited URL:
  `https://login.cribl.cloud/u/login/?state=<long_string_of_characters>`

Configure Groups

In your IDP (identity provider), configure user groups that match Permissions for both Cribl.Cloud Organizations and individual products.

Using Okta as an example, you'd map Okta groups (right side) to Cribl.Cloud Roles (left side) as follows.

IDP Group Naming

The names you define for groups in your IDP must include `Cribl` and the Role name (`Owner`, `Admin`, or `User`). They can include `Organization` or product name (`Stream`, `Edge`, or `Search`). If they don't contain a product name, the group is treated as an Organization name.

Even though IDP group names without Organization will be treated as Organization-level permissions, we recommend keeping Organization in the name for clarity.

You can use either the open or the closed format (with or without spaces) in group names. You can freely add prefixes or suffixes to Group Names that follow the formats in the examples above. Cribl will ignore these additions when mapping IDP groups to Cribl Roles. Examples:

- `SOME-LABEL-12345-CriblOrganizationOwner`
- `CriblOrganizationEditor-420`

Cribl.Cloud Role/Permission	IDP Group Name
Owner	Cribl Organization Owner /or/ CriblOrganizationOwner
Admin	Cribl Organization Admin /or/ CriblOrganizationAdmin
User	Cribl Organization User /or/ CriblOrganizationUser
Editor	**(Deprecated, will be mapped to Admin)** Cribl Organization Editor /or/ CriblOrganizationEditor
Read Only	**(Deprecated, will be mapped to User)** Cribl Organization Read Only /or/ CriblOrganizationReadOnly

Considerations for Microsoft (Azure) AD

For the groups claim configuration, you must select `Groups assigned to the application` for association. The source attribute must be set to `sAMAccountName`. Enable **Emit group name for cloud-only groups** to also return the `sAMAccountName` attribute for cloud-only groups. See the Microsoft Entra ID + OpenID Configuration topic.

Default Product Permissions

When you map external users to your Cribl Organization, their initial product-level Permissions follow a different inheritance pattern than Members configured within Cribl. This is to avoid downgrading product-

level Permissions that Organization-level `Users` might already have.

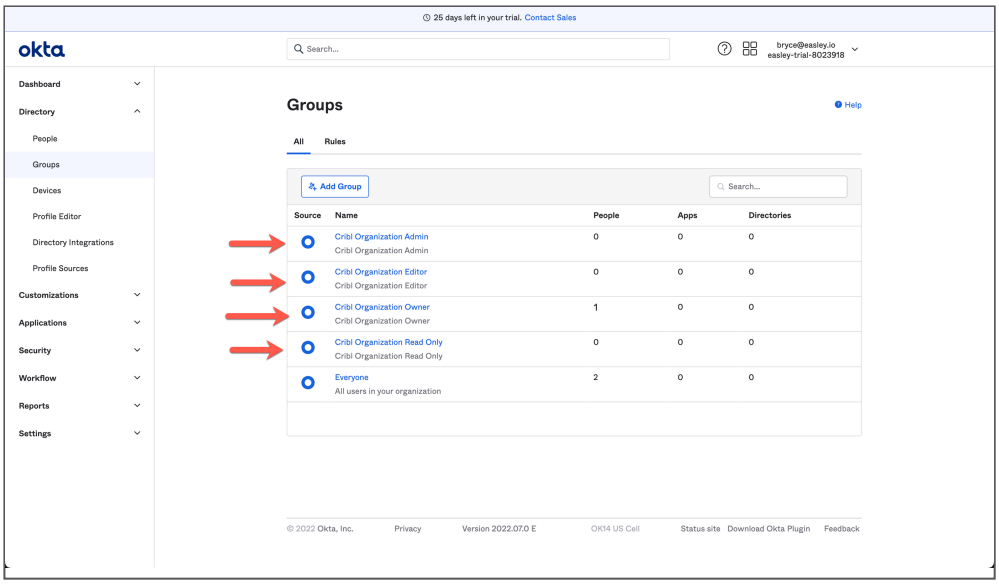
The defaults for mapped users are:

- Organization `Owner` or `Admin` inherits `Admin` Permission on all products.
- Organization `User` inherits `User` Permission on all products.
- Organization `Editor` (deprecated) inherited `editor` legacy Roles on all products.
- Organization `Read Only` (deprecated) inherited `Read Only` Permission on Stream and Edge, and `User` Permission on Search.

Group Configuration Better Practices

- A Cribl.Cloud Organization's Owner Role can be shared and transferred among multiple users. This facilitates gradual ownership transfers, corporate reorganizations, and other scenarios.
- Those users should be in both the Owner and Admin groups in your IDP. (This enables them to acquire all needed permissions across Cribl's two corresponding Roles.)
- Aside from dual-assigning the Owners, you should assign every other user only one group in your IDP. (Cribl's Admin and Editor Roles include all the permissions of the Roles below them.)

Here's an example of how groups configuration (at an early stage) might look in Okta's UI:



Mapping Okta groups to Cribl.Cloud Roles

2.6.3. OIDC/Okta Setup Example

This page expands on the overview for OIDC, offering a detailed walkthrough with Okta as the example IDP.

Cribl.Cloud supports only OIDC backchannel authentication, not front-channel.

Create OIDC App Integration

To create your app integration within Okta, navigate to the **Applications** section of your Okta environment and click **Create App Integration**.

Next, configure the app integration with the options below.

- **Sign-in method**: `OIDC - OpenID Connect`
- **Application type**: `Web Application`

General Settings

Configure the app integration's **General Settings** with the options below.

- **App integration name**: `Cribl.Cloud (OIDC)`
- **Grant type**: Select `Authorization Code` and `Refresh Token`.
- **Refresh token behavior**: Select `Use persistent token`.
- **Sign-in redirect URIs**:
  - https://login.cribl.cloud/login/callback
  - `https://manage.cribl.cloud/<organizationID>/organization/sso`
- **Sign-out redirect URIs**: https://login.cribl.cloud/v2/logout

If your IDP is PingOne, you must also configure this (non-Okta) option:

- **Authentication options**: `Allow Client Secret`

Assignments

Configure the **Assignments** pane with the following options:

- **Controlled access**: `Limited access to selected groups`
- **Selected groups**: The groups you mapped in Configure Groups

Sign-On Tab

In the **OpenID Connect ID Token** section, set the **Groups claim filter** to: `groups` : `Starts with` : `Cribl`.

To obtain the **Issuer URL** you'll need to provide to Cribl in the next section, change the value in the **Issuer** field from `Dynamic` to `Okta URL`.

This step concludes the setup procedure for Okta (or other IDP).

Submit Your App Info to Cribl

Next, provide Cribl essential details about your application, to implement SSO setup on the Cribl side.

On your Cribl.Cloud portal's Organization page > **SSO** tab, select the **OIDC** lower tab.

The **Web Application Settings** are prefilled for you, so you only need to fill in the **Cribl Cloud SSO Settings** section with the following details from your IDP client configuration:

- Client ID
- Client Secret
- Issuer URL. Copy this value from the **API** > **Settings** section of your Okta environment.

OIDC/Okta Chiclet Setup (Optional)

If you want to initiate login from your Okta instance with OIDC authentication configured, an Okta admin can configure an app integration as follows:

1. From Okta's left nav, select the **Applications** page.
2. Find the OIDC application created earlier in the OIDC/Okta Setup Example.
3. Click that application, and select **General Settings** > **Edit**.
4. In the **Initiate login URI** field, enter `https://manage.cribl.cloud/login?connection=<organizationID>` (where `<organizationID>` is your Cribl.Cloud Organization's ID).
5. Click **Save** to complete the chiclet.

Link Existing Users

To ensure that your Cribl.Cloud Organization's local users have a smooth transition to SSO, see Final SSO Steps & Troubleshooting.

2.6.4. SAML/Azure AD Setup Examples

This example uses Azure Active Directory as the identity provider (IDP).

Get URL and ID from Cribl

Cribl's terminology corresponds to Azure AD's terminology as follows:

Cribl.Cloud	Microsoft Entra ID
Single Sign-On URL	Reply URL (Assertion Consumer Service URL)
Audience URI	Identifier (Entity ID)

Create an Enterprise Application

In Microsoft Entra ID:

1. Select **Enterprise applications** (on the left) > **New application** > **Create your own application**.
2. Name your new app `Cribl.Cloud` (or any name you prefer).
3. Select **Integrate any other application you don't find in the gallery (Non-gallery)**.
4. Click **Create**.

Assign Groups

From Microsoft Entra ID's left nav:

1. Select **Users and groups**.
2. Select **Add user/group**.
3. Add the Cribl groups you created in Configure Groups.
4. Click **Assign** after selecting Groups.

Configure Single Sign-On

From Microsoft Entra ID's left nav, select **Single sign-on** > **SAML** to open the **Basic SAML Configuration** page. Then, as shown in the screenshot below:

1. Select **Add identifier** and enter the **Audience URI** value from Cribl.Cloud's SAML setup page.
2. Select **Add reply URL** and enter the two **Single Sign-on URL** values from Cribl.Cloud's SAML setup page.
3. Of these two URLs, identify the one with the `connection` query parameter, and check the checkbox to make it the **Default**.

SSO information for configuring Azure AD

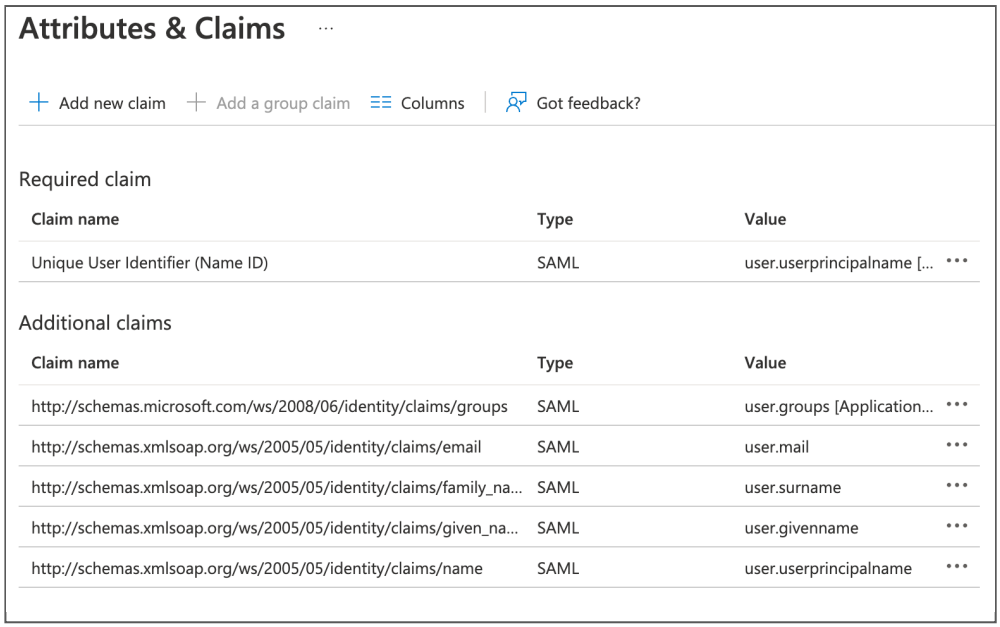
Configure Attributes and Groups Claims

In Microsoft Entra ID, edit **Attribute & Claims** as follows. Start with the claim names:

1. Change `surname` to `family_name`.
2. Change `emailaddress` to `email`.
3. Change `givenname` to `given_name`.

Next, add a group claim:

1. Select **Groups assigned to the application**.
2. As the **Source Attribute**, select: `Cloud-only group display names (Preview)`.
3. Accept the defaults for everything else, and save the new settings.



SAML Microsoft Entra ID Attribute

Submit Your App Info to Cribl

After you've created the SAML app integration in your IDP, provide Cribl essential metadata about your application, to implement SSO setup on the Cribl side.

1. On your Cribl.Cloud portal's Organization page > **SSO** tab, select the **SAML** lower tab.
2. Set the **IDP Login/Logout URL** to your Azure AD's **Set up CloudSAML** section > **Login URL** value.
3. Set the **IDP issuer** to your Azure AD's **Set up CloudSAML** section > **Azure AD Identifier** value.
4. To set the **X.509 certificate (base64-encoded)**, navigate to Azure AD's **SAML Certificates** section and download your **Base64 Certificate**.
5. Click **Test Connection**.
6. When you've verified the connection, click **Save** to complete your submission.

SAML/Azure AD Setup with My Apps Chiclet (Optional)

If you want to log into Cribl.Cloud via the Microsoft My Apps chiclet, complete the following procedure:

1. In Microsoft Entra ID, navigate to the enterprise application that you created to integrate SSO.
2. From the left nav, select **Single Sign-on**.
3. In the Enterprise Application's **Basic SAML Configurations** UI, click **Edit**.
4. In the **Sign on URL (Optional)** section, enter the following URL:

```
https://portal.cribl.cloud/login?connection=<organizationID>
```

You also need to allow self-service access to the Cribl App, or assign AD groups permissions to access the application.

Link Existing Users

To ensure that your Cribl.Cloud Organization's local users have a smooth transition to SSO, see Final SSO Steps & Troubleshooting.

2.6.5. SAML/Okta Setup Examples

This example uses Okta as the identity provider (IDP).

Get URL and ID from Cribl

Cribl will provide the following information about your Cribl.Cloud Organization, to include in the SAML application that you create in your IDP:

1. Assertion Consumer Service URL.
   Okta calls this a "Single sign-on URL," and this is the first of two URLs that your Cribl Organization's **SSO** > **SAML** tab lists under the same label. Example:
   `https://login.cribl.cloud/login/callback?connection=<$organizationID>`
2. Test SSO URL.
   Required to test your connection. Okta accepts this under "Other Requestable SSO URLs," and this is the second URL that your **SSO** > **SAML** tab lists under "Single sign-on URL." Example:
   `https://manage.cribl.cloud/api/assert`
3. Entity ID.
   Okta calls this an "Audience URI (SP Entity ID)," and your **SSO** > **SAML** tab calls it just an "Audience URI." Example: `urn:auth0:cribl-cloud-prod:<$organizationID>`

Create SAML 2.0 App Integration

1. To create your app integration within Okta, navigate to the **Applications** section of your Okta environment and click **Create App Integration**.
2. Next, create the app integration with **Sign-in method**: `SAML 2.0`.

Configure SAML Settings

1. In the app integration **SAML Settings** section, configure the following options.

- **Single sign-on URL (Assertion Consumer Service URL)**:
  `https://login.cribl.cloud/login/callback?connection=<$organizationID>`
- **Audience URI (SP Entity ID)**:
  `urn:auth0:cribl-cloud-prod:<$organizationID>`
- **Application username**: `Email`

The `nameidentifier` assertion in SAML responses must be the user's `Email`.

2. Next, click the **SAML Settings** section's **Show Advanced Settings** link. Then navigate down to configure a single row of **Other Requestable SSO URLs**, as follows:

- **URL**: From your Cribl.Cloud Organization's **SSO** > **SAML** tab, this is the second **Single sign-on URL**. It will be in this format: `https://manage.cribl.cloud/api/assert`
- **Index**: Set this to `0`.

Configure Attribute Statements

Configure **Attribute Statements** for these attributes, as shown below:

- `email`
- `given_name`
- `family_name`
- `groups`

Then save your app integration.

Submit Your App Info to Cribl

After you've created the SAML app integration in your IDP, provide Cribl with the essential metadata about your application to implement SSO setup on the Cribl side.

1. On your Cribl.Cloud portal's Organization page > **SSO** tab, select the **SAML** lower tab.

The **Web Application Settings** will be prefilled for you, and Cribl will also prefill the **SAML Assertion Mappings** based on the information you've registered with Cribl. So you only need to fill in the **SAML Configuration** section with the following details from your IDP client configuration:

- IDP SSO
- IDP Issuer
- X.509 Certificate

2. Return to your Okta environment and click **View SAML setup instructions**.

SAML/Okta Chiclet Setup (Optional)

If you want to initiate login from your Okta instance with SAML authentication configured, an Okta admin can configure an app integration as follows:

1. From Okta's left nav, select the **Applications** page.
2. Click **Browse App Catalog.**
3. From the resulting catalog, use the search bar to find and select the `Bookmark App` application.
4. From that application's page, click **Add Integration**.
5. On the **General settings** page, enter an **Application label** that will identify this app as supporting Cribl.Cloud login. (`Cribl.Cloud` is a good choice, but the label is arbitrary.)
6. In the **URL** field, enter `https://manage.cribl.cloud/login?connection=<organizationID>` (where `<organizationID>` is your Cribl.Cloud Organization's ID).
7. Click **Done**.
8. Click **Assign** and assign all of the Cribl.Cloud groups to the application.
9. The Cribl.Cloud chiclet should now be available for all users in the Cribl groups you've assigned.

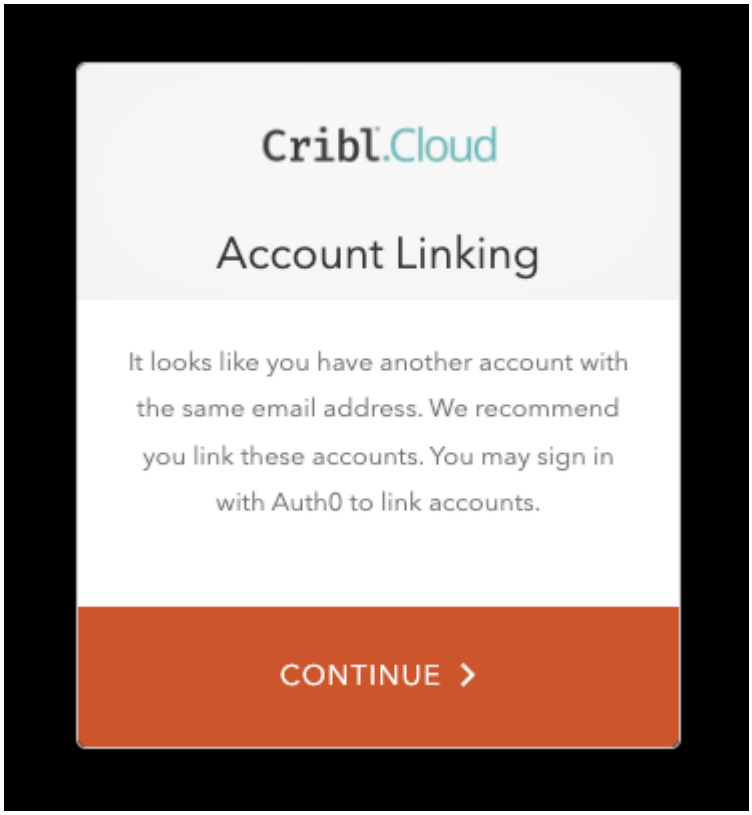
Link Existing Users

To ensure that your Cribl.Cloud Organization's local users have a smooth transition to SSO, see Final SSO Steps & Troubleshooting.

2.6.6. Final SSO Steps & Troubleshooting

Whether you're integrating with OIDC or SAML, there's one more step for users who had an existing username/password-based login on Cribl.Cloud before SSO was set up.

Upon first login with SSO, these users will see a prompt to link their identities. They should accept this prompt to ensure that their existing profile is linked with their SSO profile. (This can be a multi-step flow.)



Prompt to link accounts

Troubleshooting Resources

Cribl University offers an SSO Integration – Cribl.Cloud – Okta Troubleshooting Criblet. To follow the direct course link, first log into your Cribl University account.

To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses. Cribl's training is always free of charge.

Once logged in, check out other useful Troubleshooting Criblets and Advanced Troubleshooting short courses.

3. Deploying Edge

3.1. Deployment Planning

When you're planning an Edge deployment, consider these key factors:

- The amount of data (metrics, logs, custom command outputs, and so on) that you plan for Edge to collect from the endpoint and ingest per unit of time. For example, MB/s or GB/day.
- The amount of processing that will happen on incoming data. For example, are there a lot of transformations, regex extractions, parsing functions, field obfuscations, and so on?
- Routing and/or cloning: Is most data going to a single Destination, or is it being cloned and routed to multiple places? This is important, because Destination-specific serialization tends to be relatively expensive.
- Deploying onto servers with no internet access: If you plan to deploy Edge Nodes on air-gapped on-premises servers (servers with no internet access), you must ensure that every Edge Node can communicate with a Leader that is also on-prem (that is, not in Cribl.Cloud).

Type of Deployment

- Use Cribl Cloud to quickly launch a Cribl-hosted deployment of the combined Cribl applications suite (Edge, Stream, and Search). With this option, Cribl assumes responsibility for provisioning and managing all infrastructure, on your behalf.
- Use Single-Instance Deployment when incoming data volume is low, and/or amount of processing is light.
- Use Fleet Management to accommodate increased load. See Add/Update Edge Nodes to streamline Edge Nodes' deployment via scripting.

System Requirements

Edge Nodes should have sufficient CPU, RAM, network, and storage capacity to handle your **specific** workload. It's very important to test this before deploying to production.

Requirement Type	Requirements Details
Minimum: Edge Nodes	**OS**: Linux: RedHat, CentOS, Ubuntu, AWS Linux, Suse, Windows Server (64bit) **System**: ~1Ghz processor, 512MB RAM, 5GB of free disk space (more if persistent queuing is enabled on Edge Nodes)

Requirement Type	Requirements Details
Recommended: Leader Node	**OS**: Linux: RedHat, CentOS, Ubuntu, AWS Linux, Suse (64bit) **System**: +4 physical cores, +8GB RAM, 5GB free disk space

Edge Nodes can't be installed on Windows laptops/desktops.

Linux Requirements

- 64-bit kernel >= 3.10 and glibc >= 2.17.
- SELinux: Enforcing mode is supported, but not required.

Tested Platforms (Linux)

- **OS (Intel Processors)**:
  - Ubuntu 16.04+, Debian 9+, RHEL 7+, CentOS 7+, SUSE Linux Enterprise Server 12+, Amazon Linux 2014.03+.
- **OS (ARM64 Processors)**:
  - Ubuntu (14.04, 16.04, 18.04, and 20.04), CentOS 7.9, and Amazon Linux 2.

Browser Support

- Chrome, Firefox, Safari, and Microsoft Edge: the five most-recent versions.

Leader Scalability

Our System Requirements suffice for Leaders handling up to 3,000 Edge Nodes. To support more than 3,000 Nodes, see our Scaling Edge Beyond 3k Nodes guide.

Leader/Edge Nodes Compatibility

Leaders on v.4.2.x are compatible with Edge Nodes on v.4.1.2, v.4.1.3, and later. Due to a security update, Edge Nodes running on v.4.0.4 cannot receive configurations from v.4.2.x Leaders. For details, see Leader and Edge Nodes Compatiblity.

About Licensing

- **Free**: Entitles users to manage up to 100 Edge Nodes, in a single Fleet, and ingest up to 1TB/day.
- **Enterprise**: Entitles users to manage unlimited Edge Nodes, across unlimited Fleets and ingest up to its rated volume.

Permissions and Rights

You do **not** need elevated privileges to install Edge. However, you will need:

- Administrator (or equivalent) rights in order to enable Edge to start on boot.
- Sufficient rights to access the resources that need monitoring – files, metrics, etc.

Performance Considerations

As with most data-collection and -processing applications, Cribl Edge's expected resource utilization will be proportional to the data volume and type of processing. For instance, a Function that adds a static field on an event will perform faster than one that applies a regex to find and replace a string.

- Processing performance is proportional to CPU clock speed.
- All processing happens in-memory.
- Processing does not require significant disk allocation.

Thinking about your planned Edge deployment as a whole, it's critical to consider **cardinality**: how many Fleets, of what size, will send metrics to a given Leader. In a very low-cardinality deployment, a Leader Node might aggregate metrics from one small Fleet; in a very high-cardinality deployment, a Leader Node might aggregate metrics from many large Fleets.

Too many Edge Nodes sending too many metrics can degrade Leader performance and/or integrity. By carefully choosing what metrics each Fleet sends to its Leader Node, you can ensure that you collect the metrics most important to you, within the limits of the Leader's capacity to process them.

As part of your deployment planning, decide which of the following four sets of metrics each Fleet in your deployment will send to its Leader Node:

- **Minimal**: Limited to metrics for events and bytes in and out. Recommended for high-cardinality deployments.
- **Basic** (the default set): All the metrics required to display all monitoring data available in the Edge UI. Contains all the metrics in the **Minimal** set, and more.
- **All**: All metrics, sent without filtering.
- **Custom**: Only metrics that match a JavaScript expression that you define.

For a listing of the metrics each set provides, see Controlling Metrics in Cribl Edge.

How Edge Works with AppScope

AppScope is an Open Source utility from Cribl that can send events, metrics, and/or payloads from a process or application to an AppScope Source on Edge.

Here's what to keep in mind when you plan to use Edge and AppScope together:

- AppScope is bundled with Edge and does not need to be installed.
- Running Edge with AppScope requires that Edge either run as root on a Linux host, or in a Docker container that is in privileged mode. (If this is a problem for you, see this workaround).
- To work with an AppScope host, all Edge nodes in a Fleet should be either all in Docker containers, or, all directly on Linux hosts. Plan to keep a ratio of one AppScope Source to one Fleet.

Fleet Checklist

This section compiles basic checkpoints for successfully launching a Fleet.

1. System

- 1 Leader Node (recommended: 8 vCPU/16 GB RAM).
- N Edge Nodes (depending on your environment).
- Free license, or acquire an Enterprise/Trial License from the Cribl Sales Team.

2. Configure Leader Node

- Install `git`, if not present (`yum install git`).
- Open the necessary ports.
- Download, install, and launch Cribl Edge (Linux, Windows).
- Configure as a Leader.
- Enable Start on Boot.
- Install License.

3. Configure Edge Nodes

- Enable GUI Access. Administrators will need to connect to the TCP:9420 port on each Node.
- Download, install, and launch Cribl Edge (Linux, Windows).

- Configure as a Managed Edge Node.
  - Point to the Leader address (optionally, use the configured access token).
  - Give each Edge Node an arbitrary tag like `POV`.
- Enable Start on Boot.

4. Map Edge Nodes to Fleets

- On the Leader Node, create a Fleet.
  - Name the Fleet (abitrarily) `POV`.
- On the Leader Node, confirm that Edge Nodes are connecting.
  - From the Leader Node's top nav, click **Manage** and select **Edge Nodes**.
- Map Nodes to the `dev` Fleet.
  - Use the Filter to select the tag you applied when configuring Nodes: `cribl.tags.includes('POV')`.

Configure Connection Processes

To configure the connection processes for your Edge Nodes:

1. From the top nav, select **Settings** > **Global Settings** > **Service Processes** > **Services**.
2. Confirm that **Connections listener number of processes** is set to its default `1`.

3.1.1. Scaling Edge Beyond 3k Nodes

Learn how to scale Cribl Edge deployments to support one Leader managing up to 50,000 Nodes.

Our standard Deployment Planning guide is a great place to start if you intend to deploy 3,000 or fewer Nodes in your Edge environment.

In version 4.6 and newer, a Leader can support a maximum of 50,000 Edge Nodes. To support this higher Node limit, we've made some changes to Edge. This guide covers the changes you can expect to see in the Edge UI, some areas where you might encounter slowness as we continue working on supporting more Nodes, and the specific Leader configuration we've tested for best performance.

What to Expect in the User Interface

We've updated the UI to help visually and behaviorally support Fleets that contain more than 3,000 Nodes. This section will highlight the key changes, and alert you to potential issues you may encounter.

Searching for Nodes in the Explore Tab

Typing into the **Node to explore** field returns only the first 100 matching Nodes, sorted alphabetically by hostname.

Previewing Node Mappings

When you're previewing Node mappings from the **Manage** menu, **Mappings** submenu, the results are limited to the first 10,000 Mappings.

Viewing the Status Tab on a Source or Destination

When there are more than 50 Nodes in a Fleet, the **Status** column won't be visible in the list of Nodes (from a Source or Destination).

To view the status for an individual Node, expand it in the list. You can only expand one Node at a time. Refresh the list of Nodes and their status with the **Refresh** button.

Recommended Leader Configuration

In general, these are the settings you should configure on the Leader in order for it to be able to manage more than 3,000 Nodes:

- Deploy one CPU for every 3,000 Edge Nodes. To scale to 50,000 Nodes, you might need to deploy around 16 CPUs (depending on your specific workload).
- Configure one Connection Process for every 10,000 Edge Nodes.
- While one Edge Fleet *can* contain all 50,000 Nodes, we've seen slowdown and performance issues in testing. Do not push new configurations to more than one Fleet at a time.
- Increase memory to 32GB.
- As a starting point, increase the RAM of the Leader instance's physical host by 200MB RAM per Edge Fleet. Continue to increase memory per Edge Fleet if needed.
- In the current release, each Leader can support a maximum of 50,000 Edge Nodes.
- In **Fleet Settings**, configure internal metrics to **Minimal**. See Performance Considerations for more information.
- There is no change to the default values when pushing config bundles from the Leader.

Cribl University offers a course titled Cribl Edge Architecture: Architecture & > Sizing that provides additional best practices and guidance. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

Upgrading 50,000 Edge Nodes

We recently rolled out a new upgrading framework in which Edge Nodes asynchronously pull down upgrade packages from the Leader, rather than the Leader pushing packages to Nodes all at one time. Nodes communicate to the Leader every 60 seconds via a heartbeat.

Additionally, Nodes request upgrade packages from the Leader in batches of 500. This allows you to upgrade Nodes without overloading the Leader, even with a large number of Nodes. In our controlled test environment, we found that upgrading 50,000 Nodes can take anywhere from 25-40 minutes. This is an expected time frame to upgrade that number of Nodes.

Current Limitations

We're continuing to address places where we've noticed an altered UI experience. This is what we've identified so far:

- **Logs**: The drop-down where you can select an Edge Node will display a maximum of 100 Nodes.
- **Map view**: To improve performance, the Map view doesn't load all Nodes.

- **Import Edge Data**: The **Import Edge Data** window in Cribl Stream might respond slowly if you are attempting to view Edge sample data and you have a large number of Nodes.
- **Export list as**: It might take longer than normal to export the list when the list contains a large number of Nodes.
- **Test tab**: The **Test** tab on Destinations lists all Nodes and might take a long time to load.
- **KMS Status report**: In **Fleet Settings**, the **KMS Status report** (under **Security**) does not display all Nodes. Instead, search for your desired Node.

3.2. Installing Cribl Edge On Linux

First, download the install package, which is the same binary as Cribl Stream, to your Linux machine.

Next, ensure that required ports are available (see Network Ports).

Then, you can install and run any of the following:

- Single Cribl Edge instance.
- Cribl Edge and Cribl Stream on the same host.
- Multiple installations of the same product, on the same host.

Single Cribl Edge Instance

These instructions explain how to run Cribl Edge locally, as a single-instance deployment on your own machine.

Un-tar the install package in a directory of choice, and rename the resulting `cribl` directory as `cribl-edge`, e.g.:

```
cd /opt/
tar xvzf cribl-<version>-<build>-<arch>.tgz
mv /opt/cribl/ /opt/cribl-edge
```

Set the renamed directory, (e.g., `/opt/cribl-edge/`), as your `$CRIBL_HOME` directory.

To make the `$CRIBL_HOME` env variable available on your command line, you can:

- Assign it once, using the `export` command:
  `export CRIBL_HOME=/opt/cribl-edge`
- Set it as a default, by adding it to your to your terminal profile file.

Next, navigate to `$CRIBL_HOME/bin`. Here, you can use `./cribl` commands to:

- **Set to Edge mode**: `./cribl mode-edge`
- **Start**: `./cribl start`
- **Stop**: `./cribl stop`

- **Reload**: `./cribl reload`
- **Restart**: `./cribl restart`
- **Get status**: `./cribl status`

You can change the hostname and port, by adding `-H` (address) and `p` (port options). E.g: `./cribl mode-edge -H 0.0.0.0 -p 8123`

Next, go to `http://localhost:9420` and log in with default credentials (`admin:admin`). You can now start configuring Cribl Edge with Sources and Destinations, or start creating Routes and Pipelines.

Cribl Edge and Cribl Stream on the Same Host

You can run an Edge Node on a Cribl Stream Leader Node, or an Edge Node and a Stream Worker Node on the same host. To accommodate these scenarios, each product has a distinct service name:

- `cribl-edge.service` for Cribl Edge.
- `cribl.service` for Cribl Stream.

Here, Cribl recommends un-tarring the download package twice, into two separate directories. This setup frees you to update and run each product individually. You could choose, e.g., `/opt/cribl-edge` and `/opt/cribl`:

```
cd /opt
tar zxvf /tmp/cribl-<version>-<build>-<arch>.tgz
mv cribl cribl-edge
tar zxvf /tmp/cribl-<version>-<build>-<arch>.tgz
```

The installation and configuration sequence will be the same for each product:

1. Change the ownership for both installations.
   Run Cribl Stream as a non-privileged user:

   ```
   chown -R cribl:cribl /opt/cribl
   ```

   Run Cribl Edge as a non-privileged user:

   ```
   chown -R cribl:cribl /opt/cribl-edge
   ```

Do NOT Run Cribl Edge as Root!

To listen on low ports 1-1024, Cribl Edge needs privileged access. You can enable this on systemd by adding this configuration key to your `override.conf` file:

```
[Service]
AmbientCapabilities=CAP_NET_BIND_SERVICE
```

If you want to add extra capabilities, such as reading certain resources (e.g., `/var/log/*`), add `CAP_DAC_READ_SEARCH` in a space-separated format as follows:

```
[Service]
AmbientCapabilities=CAP_NET_BIND_SERVICE CAP_DAC_READ_SEARCH
```

Alternatively, you can use ACLs to allow Cribl Edge to read files.

For details, see Running Edge as an Unprivileged User.

2. Set the correct mode, and configure each installation as a service. The `-H` and `-p` parameters are required.
   For Cribl Edge:

```
/opt/cribl-edge/bin/cribl mode-managed-edge -H <leader-hostname-or-IP> -p <por

/opt/cribl-edge/bin/cribl boot-start enable
```

   If you are setting up Cribl Edge in single-instance mode, make sure to set `mode-edge` instead of `mode-managed-edge` .
   For Cribl Stream:

```
/opt/cribl/bin/cribl mode-worker -H <leader-hostname-or-IP> -p <port> [options

/opt/cribl/bin/cribl boot-start enable
```

Managing IP Addresses

By default, Cribl Edge's API listens on port `9420`, instead of on Cribl Stream's default `9000` port. Some things to note:

- You can set the `CRIBL_EDGE` environment variable to any value to bind to `0.0.0.0`, instead of to the `127.0.0.1` address.
- If you are starting Cribl Edge from the CLI, make sure you set the `-H` parameter to `0.0.0.0`.

- If you connect your Edge Node to a Leader, the Node will automatically update the binding IP address to the one configured in the Leader's Fleet Settings. For an Edge Node to always listen on all IP addresses, you must update the Leader's **Host** value. (In the Leader's UI, go to **Fleet Settings** > **System** > **General Settings**, and change the **Host** field to `0.0.0.0`.)

Continue with Setting Up Leader and Edge Nodes for Edge, and Setting Up Leader and Worker Nodes for Stream.

Cribl University offers a course titled How to Install & Configure Edge on Linux that illustrates the installation steps. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

Multiple Installations of Same Product on Same Host

There are situations where it makes sense to run multiple Cribl Edge or Cribl Stream installations on the same host. For example, suppose two departments want to collect the data at the edge – and each wants to process the data differently, and to deploy its own Helm chart as a daemonset.

To support this: After un-tarring the installation package, copy or move it into a separate directory for each installation. For example, if you're creating two Cribl Edge installations:

```
cd /opt
tar zxvf /tmp/cribl-<version>-<build>-<arch>.tgz
cp -r cribl/ cribl-edge-01/
mv cribl/ cribl-edge-02/
```

The installation and configuration procedure is the same as described for collocated Cribl Edge and Cribl Stream above, except that you must:

- Omit steps pertaining to the product you are **not** installing.
- Repeat steps pertaining to the product you are **are** installing – once for each instance of the product.

Ensure that each instance of the product runs on its own dedicated port. Either:

- Specify different ports when you run the `mode-managed-edge` or `mode-worker` command; or,

- On the host, set the `CRIBL_AUTO_PORTS` environment variable to `1`.

Troubleshooting Resources

Cribl University's Troubleshooting Criblet on Switching from Cribl Edge to Cribl Stream demonstrates these techniques for integrating Edge with Stream. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Troubleshooting Criblets and Advanced Troubleshooting short courses.

3.2.1. Enabling Start On Boot

Cribl Edge ships with a CLI utility that can update your system's configuration to start Edge at system boot time. The basic format to invoke this utility is:

```
[sudo] $CRIBL_HOME/bin/cribl boot-start [enable|disable] [options] [args]
```

You must run this command as root, or with `sudo`. For options and arguments, see the CLI Reference.

The script will create a user named `cribl` to install, own, and run Cribl Edge/Stream.

Most, if not all, popular Linux distributions use `systemd` now to start processes at boot, while older or more obscure distributions may still use `initd`. Verify with your Linux distribution vendor if you aren't sure which method your systems use in order to know which procedure listed below to follow.

Using systemd

To **enable** Cribl Edge to start at boot time with **systemd**, you must run the `boot-start` command. Make sure you first create any user you want to specify to run Edge. E.g., to run Edge on boot as existing user `cribl`, you'd use:

```
sudo $CRIBL_HOME/bin/cribl boot-start enable -m systemd -u cribl
```

This will install a unit file (as shown below) named `cribl-edge.service`, and will start Cribl Edge at boot time as user `cribl`. A `-configDir` option can be used to specify where to install the unit file. If not specified, this location defaults to `/etc/systemd/system/`.

If necessary, change ownership for the Cribl Edge installation:

```
[sudo] chown -R cribl $CRIBL_HOME
```

Next, use the `enable` command to ensure that the service starts on system boot:

```
[sudo] systemctl enable cribl-edge
```

To **disable** starting at boot time, run the following command:

```
sudo $CRIBL_HOME/bin/cribl boot-start disable
```

Other available `systemctl` commands are:

```
systemctl [start|stop|restart|status] cribl-edge
```

Note the file's default `65536` hard limit on maximum open file descriptors (known as a `ulimit`). The minimum recommended value is `65536`. Linux tracks this per user account. You can view the current soft `ulimit` for max open file descriptors with `$ ulimit -n` while logged in as the same user running the `cribl` binary.

Installed systemd File

```
[Unit]
Description=Systemd service file for Cribl Edge.
After=network.target

[Service]
Type=forking
User=cribl
Restart=always
RestartSec=5
LimitNOFILE=65536
PIDFile=/install/path/to/cribl/pid/cribl.pid
ExecStart=/install/path/to/cribl/bin/cribl start
ExecStop=/install/path/to/cribl/bin/cribl stop
ExecReload=/install/path/to/cribl/bin/cribl reload
TimeoutSec=60

[Install]
WantedBy=multi-user.target
```

Persisting Overrides

By default, disabling and re-enabling boot start will regenerate the `cribl-edge.service` file. To persist any overrides – such as proxy or privileged port usage – use this command:

```
systemctl edit cribl-edge
```

This opens a text editor that prompts you to enter overrides, then saves them to a persistent file at:

```
/etc/systemd/system/cribl-edge.service.d/override.conf
```

Do NOT Run Cribl Edge as Root!

> To listen on low ports 1-1024, Cribl Edge needs privileged access. You can enable this on systemd by adding this configuration key to your `override.conf` file:
>
> ```
> [Service]
> AmbientCapabilities=CAP_NET_BIND_SERVICE
> ```
>
> If you want to add extra capabilities, such as reading certain resources (e.g., `/var/log/*`), add `CAP_DAC_READ_SEARCH` in a space-separated format as follows:
>
> ```
> [Service]
> AmbientCapabilities=CAP_NET_BIND_SERVICE CAP_DAC_READ_SEARCH
> ```
>
> Alternatively, you can use ACLs to Allow Cribl Edge to Read Files.
>
> For details, see Running Edge as an Unprivileged User.

Using initd

To **enable** Cribl Edge to start at boot time with **initd**, you must run the `boot-start` command. If the user that you want to run Cribl Edge does not exist, create that user prior to executing. E.g., running Edge as user `cribl` on boot:

```
sudo $CRIBL_HOME/bin/cribl boot-start enable -m initd -u cribl
```

This will install an `init.d` script in `/etc/init.d/cribl.init.d`, and will start Cribl Edge at boot time as user `cribl`. A `-configDir` option can be used to specify where to install the script. If not specified, this location defaults to `/etc/init.d`.

If necessary, change ownership for the Cribl Edge installation:

```
[sudo] chown -R cribl $CRIBL_HOME
```

To **disable** starting at boot time, run the following command:

```
sudo $CRIBL_HOME/bin/cribl boot-start disable
```

To control Cribl Edge, you can use the following initd commands:

```
service cribl-edge [start|stop|restart|status]
```

Persisting Overrides on initd

Notes on preserving required permissions across restarts and upgrades:

To read file resources on a Linux system that are typically restricted to the root user, you can add the `CAP_DAC_READ_SEARCH` capability. For example:

On some OS versions (such as CentOS), you must add an `-i` switch to the `setcap` command. For example: `# setcap -i cap_dac_read_search=+ep $CRIBL_HOME/bin/cribl`

Important: Upgrading Edge will remove the `CAP_DAC_READ_SEARCH` capability from the `cribl` executable, so you'll need to re-run the appropriate `setcap` command after each upgrade.

3.2.2. Running Edge As An Unprivileged User

Privileged access might be necessary if Cribl Edge needs to read certain resources (e.g., `/var/log/*`), or to listen on low ports 1–1024. Features like auto-discovery of logs and information in the Processes UI also require permissions to access `/proc`. The regular non-root permissions are not sufficient in these cases.

There are two alternatives to running Cribl Edge as `root`:

- Set Linux capabilities that grant Cribl Edge sufficient rights to perform specific privileged tasks. For details, see Set Capabilities for Cribl Edge below.
- Take advantage of Linux systems' option to layer an Access Control List (ACL) over the default Linux permissions. By using ACLs, you can assign a more specific set of permissions to a file or directory without (necessarily) changing the base ownership. For details, see Using ACLs to Allow Cribl Edge to Read Files.

Set Capabilities for Cribl Edge

Capabilities are permissions that grant privileged processes sufficient rights to accomplish a specific task, based on a kernel privilege. To run Cribl Edge as non-root user, consider setting the following capabilities:

Capability	Permissions
`CAP_NET_BIND_SERVICE`	Allows Cribl Edge to push Sources that bind to TCP/UDP port numbers below `1024`.
`CAP_DAC_READ_SEARCH`	Allows the `cribl` user to access files in **Explore** > **Files** > Manual/Browse, and to access the File Monitor Source's Manual mode feature. This capability bypasses the default Linux permissions for files and directories.
`CAP_SYS_PTRACE`	Allows the `cribl` user to scan open files for running processes, to discover active logs in **Explore** > **Files** > Auto, and to access the File Monitor Source's Auto-mode feature.

For details about setting these capabilities, see Persisting Overrides.

OS-Specific Options

To read file resources on a Linux system that are typically restricted to the root user, you can add the `CAP_DAC_READ_SEARCH` capability. For example:

On some OS versions (such as CentOS), you must add an `-i` switch to the `setcap` command. For example:

```
# setcap -i cap_dac_read_search=+ep $CRIBL_HOME/bin/cribl
```

Upgrading Edge will remove the `CAP_DAC_READ_SEARCH` capability from the `cribl` executable, so you'll need to re-run the appropriate `setcap` command after each upgrade.

Fallbacks from Privileged Access

If installing and running Edge with `root`-level privileges is forbidden or impractical in your environment, certain Sources, like Exec and System Metrics, can run on a user with lower permissions. You can also run the File Monitor Source in `Manual` mode, and collect from any files that this user can read.

Privileges, Edge, and AppScope

When you need to run Edge as an unprivileged user, you can use the AppScope CLI to send data to Edge (instead of "driving" AppScope from the Edge UI). With this workaround, the unpriviledged Edge can still:

- Monitor status of scoped processes
- Receive data from scoped processes
- Change the configuration of scoped processes

See the AppScope CLI documentation.

3.2.3. Installing Edge On Linux Via RPM

If you need to install Edge in a tightly managed Linux environment, use the signed RPM package installation method.

The RPM will create a `cribl-edge` user and a `cribl-edge` group. The `cribl-edge` user will own the `cribl-edge` configuration file (stored in the `/etc/sysconfig` directory). The `config` file contains the Leader's auth token. To keep the Leader's auth token secure, only `root` and the `cribl-edge` user can read/modify the `config` file contents.

RPM Download Links

Edge Version	RPM Download Link	GPG Key
4.5	• x64 • arm64	4.5.0 RPM Public Key
4.5.1	• x64 • arm64	4.5.1 RPM Public Key

Verifying the RPM Package Signature

To verify the RPM package signature, first enable local GPG package checking. Next, download and import the GPG key and verify its signature.

Enable Local GPG Package Checking

Your environment may not have the GPG check for local packages enabled by default.

To enable the local package GPG checks for `yum` or `dnf`, add `localpkg_gpgceheck=1` to the `main` stanza in:

- `/etc/yum/conf` for `yum`
- `/etc/dnf/dnf.conf` for `dnf`

For example:

```
[main]
...
localpkg_gpgcheck=1
```

Download and Import the GPG Key

Cribl's public GPG key ensures the authenticity and integrity of the Cribl RPM package.

1. Download the GPG key: https://cdn.cribl.io/dl/CRIBL_RPM_PUBLIC_GPG.
2. Manually import the public GPG key to `rpm` using this command:

   ```
   rpm --import <URL/filepath>
   ```

   For example:

   ```
   rpm —-import https://cdn.cribl.io/dl/4.5.0/CRIBL_RPM_PUBLIC_GPG
   ```

3. Verify the package signature using this command:

   ```
   rpm -Kv <rpm-file>
   ```

   This command ensures the package's integrity and authenticity, returns information about the package, and checks if the package has been tampered with since being signed by the original developer or distributor. This helps prevent you from installing potentially malicious or compromised packages.

Installing Edge via RPM

Follow the installation steps to install the RPM package on a Cribl Linux Edge Node.

Out of the box, the RPM will install the software into the `/opt` directory. The RPM package will configure but not enable the Systemd service. You will need to enable the software service separately.

To install Cribl Edge via RPM:

1. Open a CLI and run one of the commands, depending on the processor:

```
sudo yum install https://cdn.cribl.io/dl/4.5.0/cribl-edge-4.5.0-1a628515-linux
```

Or

```
sudo yum install https://cdn.cribl.io/dl/4.5.0/cribl-edge-4.5.0-1a628515-linux
```

This command does the following:

- Creates the `cribl-edge` user and group, if it doesn't exist.
- Puts the contents of Cribl's TGZ into `/opt/cribl-edge` and makes `root/root` the owner.
- Creates the `cribl-edge` systemd service, which is disabled by default.
- Creates `/etc/sysconfig/cribl-edge.conf` owned by `cribl-edge:cribl-edge` and only accessible by the `cribl-edge` user.
- Creates `/var/lib/cribl-edge/` owned by `cribl-edge:cribl-edge` and only accessible by the `cribl-edge` user.
- Sets `CRIBL_INSTALL_TYPE` to `RPM` in the service destination, to block distributed upgrades, as upgrading from the Leader bypasses the security of RPM.
- Sets `CRIBL_VOLUME_DIR` to `/var/lib/cribl-edge` in the service destination, so configs and logs are saved to the `cribl-edge` directory.

2. Edit the environment variables in `/etc/sysconfig/cribl.conf`, in order for Cribl Edge to connect to a Leader. The RPM installs `cribl-edge` in standalone mode by default (`CRIBL_DIST_MODE = edge`).

   To deploy Edge in a managed deployment, edit these environment variables:

   - `CRIBL_DIST_MODE = managed-edge`
   - `CRIBL_DIST_MASTER_URL=tls://<authToken>@leader:4200`

If you're trying to change the mode to `managed-edge`, you'll need to install Git separately if it is not already installed.

3. Enable the service:

- `systemctl enable cribl-edge` to enable the service on boot.
- `systemctl start cribl-edge` to start it manually.

Upgrading Edge via RPM

To upgrade Cribl Edge installed via RPM, you must use the command line:

```
sudo yum update <CDN URL for new RPM>
```

We've disabled UI upgrades, because upgrading from the Leader bypasses the security of RPM.

Any changes you make to environment variables, like editing or adding new variables, are maintained when you upgrade, as well as when the service starts (on boot or manually).

Uninstalling Edge via RPM

To remove Cribl that was installed via RPM, use the command line:

```
sudo yum remove cribl-edge
```

Uninstalling Cribl removes the `cribl` or `cribl-edge` service, the `cribl` binary from `/opt/`, and the service file from `/lib/systemd/system`.

Configuration files are maintained, along with any environment variables you edited when you first installed Cribl. This means you can pull in your previous configuration(s) if you decide to reinstall Cribl via RPM.

The `cribl-edge` directory is also maintained, along with the data, logs, and defaults of your configuration.

3.2.4. Anti-Virus Exceptions

If you are running anti-virus software on a Edge instance's host OS, here are general guidelines for minimizing accidental blockage of Edge's normal operation.

Your overall goals are to prevent the anti-virus software from locking any files while Edge needs to write to them, and from triggering any changes that Edge would detect as needing to be committed.

First, if Persistent Queues are enabled on any Destinations, exclude any directories that these Destinations write to. This is especially relevant if you're writing queues to any custom locations outside of `$CRIBL_HOME`.

Next, for any non-streaming Destinations that you've configured, exclude their staging paths.

Next, exclude these subdirectories of `$CRIBL_HOME`:

- `state/`
- `log/`
- `.git/` (usually only exists on Leader Nodes)
- `groups/` (on Leader Nodes)
- `local/` (on Edge or Leader)

Finally, avoid scanning any processes. Except for the queueing/staging directories already listed above, Edge runs everything in memory, so scanning process memory will slow down Edge's processing and reduce throughput.

3.2.5. System Proxy Configuration On Linux

You can direct all outbound HTTP/S requests to go through proxy servers. You do so by setting a few environment variables before starting Cribl Edge, as follows:

Configure the `HTTP_PROXY` and `HTTPS_PROXY` environment variables, either with your proxy's IP address, or with a DNS name that resolves to that IP address. Optionally, follow either convention with a colon and the port number to which you want to send queries.

`HTTP_PROXY` examples:

```
$ export HTTP_PROXY=http://10.15.20.25:1234
$ export HTTP_PROXY=http://proxy.example.com:1234
```

`HTTPS_PROXY` examples:

```
$ export HTTPS_PROXY=http://10.15.20.25:5678
$ export HTTPS_PROXY=http://proxy.example.com:5678
```

In the above examples, note that when you set an `HTTPS_PROXY` environment variable, the referenced URL should generally be in `http` format.

Restarts and Case Conflicts

Initial configuration of, and changes to, these variables require restarting Cribl Edge on the affected Nodes, if the application is already running when you apply the changes.

The environment variables' names can be either uppercase or lowercase. However, if you set duplicate versions of the same name, the lowercase version takes precedence. E.g., if you've set both `HTTPS_PROXY` and `https_proxy`, the IP address specified in `https_proxy` will take effect.

HTTP and/or HTTPS?

Several Cribl Edge endpoints rely on the HTTPS protocol – the Cribl telemetry endpoint, which must be accessed with some license types, as well as the CDN used to propagate application updates and certain documentation features (API Reference and docs PDFs).

You might configure certain other Cribl Edge features (such as REST API Collectors) that require access to HTTP endpoints. For maximum flexibility, consider setting environment variables to handle both the HTTPS and HTTP protocols.

Proxy Configuration with systemd

If you are proxying outbound traffic and starting Cribl Edge using systemd, add your proxy environment variables to the systemd override file (see Persisting Overrides). Add statements of this form:

Installed systemd File

```
[Service]
Environment=http_proxy=<yourproxy>
Environment=https_proxy=<yourproxy>
Environment=no_proxy=<no_proxy_list>
```

This will prevent Cribl Edge from throwing "failed to send anonymized telemetry metadata" errors.

Authenticating on Proxies

You can use HTTP Basic authentication on HTTP or HTTPS proxies. Specify the user name and password in the proxy URL. For example:

```
$ export HTTP_PROXY=http://username:password@proxy.example.com:1234
$ export HTTPS_PROXY=http://username:password@proxy.example.com:5678
```

Bypassing Proxies with `NO_PROXY`

If you've set the above environment variables, you can negate them for specified (or all) hosts. Set the `NO_PROXY` environment variable to identify URLs that should bypass the proxy server, to instead be sent as direct requests. Use the following format:

```
$ export NO_PROXY="<list of hosts/domains>"
```

Cribl recommends including the Leader Node's host name in the `NO_PROXY` list.

`NO_PROXY` Usage

Within the `NO_PROXY` list, separate the host/domain names with commas or spaces.

Optionally, you can follow each host/domain entry with a port. If not specified, the protocol's default port is assumed.

To match subdomains, you must either list them all in full (for example, `NO_PROXY=foo.example.com,bar.example.com`), or apply a wildcard by prefixing the domain name with a period or `*.`: `NO-PROXY=".example.com` or `NO-PROXY="*.example.com`.

To match the whole domain including its subdomains, add it both with and without wildcard to the list: `NO-PROXY="example.com,.example.com`.

To disable all proxies, use the `*` wildcard: `NO_PROXY="*"`. `NO_PROXY` with an empty list disables no proxies.

Cloud `NO_PROXY` Usage

You must include any cloud metadata endpoints (such as the AWS Instance Metadata Service) in the `NO_PROXY` list:

- AWS EC2 and Azure VM instances must include `169.254.169.254` in the list. If using IPv6 on AWS EC2, add `fd00:ec2::254` to the list.
- AWS ECS Fargate tasks must include `169.254.170.2`.
- GCP (Google Cloud Platform) VM instances must include `metadata.google.internal` and `169.254.169.254`.

Where Proxies Apply

Proxy configuration is relevant to the following Cribl Edge components that make outbound HTTP/S requests:

Destinations

- S3 Compatible Stores
- AWS Kinesis Streams
- AWS CloudWatch Logs
- AWS SQS
- Azure Blob Storage
- Azure Monitor Logs
- Cribl HTTP

- CrowdStrike Falcon LogScale
- Elasticsearch
- Grafana Cloud
- Honeycomb
- Loki
- Prometheus
- Splunk HEC
- Webhook

Notification Targets

- PagerDuty
- Webhook

Testing Proxies

To initially test your proxy configuration, consider setting up a simple, free proxy server like mitmproxy (https://mitmproxy.org/), and then monitoring traffic through that server. Verify that you can trace proxied requests from your Cribl Edge instance, and can validate that outgoing requests (to Destinations) are working properly.

Proxying Multiple Edge Instances in One Browser

Cribl Edge stores authentication tokens based on each http header's URI scheme, host, and port information. Within a given browser, Cribl Edge enforces a same-origin policy to isolate instances.

This means that if you want to run multiple proxied Cribl Edge instances in one browser session, you must assign them different URI schemes, hosts, and/or ports. Otherwise, logging into an extra Cribl Edge instance will expire the prior instance's session and log it out.

For example, assume that you've set up this pair of Apache proxy forward rules:

- https://web/cribla forwards to `cribl_hosta:8001/cribla`.
- https://web/criblb forwards to `cribl_hostb:8001/criblb`.

These two proxied addresses cannot be run simultaneously in the same browser session. However, this pair – which lead with separate URI schemes – could:

- https://web/cribla forwards to `cribl_hosta:8001/cribla`.

- https://web2/criblb forwards to `cribl_hostb:8001/criblb`.

Where separate instances **must** share URI formats, a workaround is to open the second instance in an incognito/private browsing window, or in a completely different browser.

3.3. Installing Cribl Edge On Windows

You can install Cribl Edge on Windows Server 2016, 2019, or 2022. To start:

1. Set a compatible browser as your default browser: Microsoft Edge, Firefox, or Chrome. (We support the five most-recent versions of these browsers. Cribl Edge is not compatible with Internet Explorer.)
2. Ensure that the required ports are available (see Network Ports).
3. Go to the Cribl Download page and set the **Software** drop-down to `Cribl Edge for Windows`.
4. Click **Download Now** to get the `.msi` installer.

You can also concatenate and copy/paste the bootstrap script in your command prompt to add Windows Node.

Select the Installation Type

You now have two installation options:

- Launch the `.msi` to install via an interactive wizard.
- Use the `.msi` to install via a command prompt, or to script bulk installs.

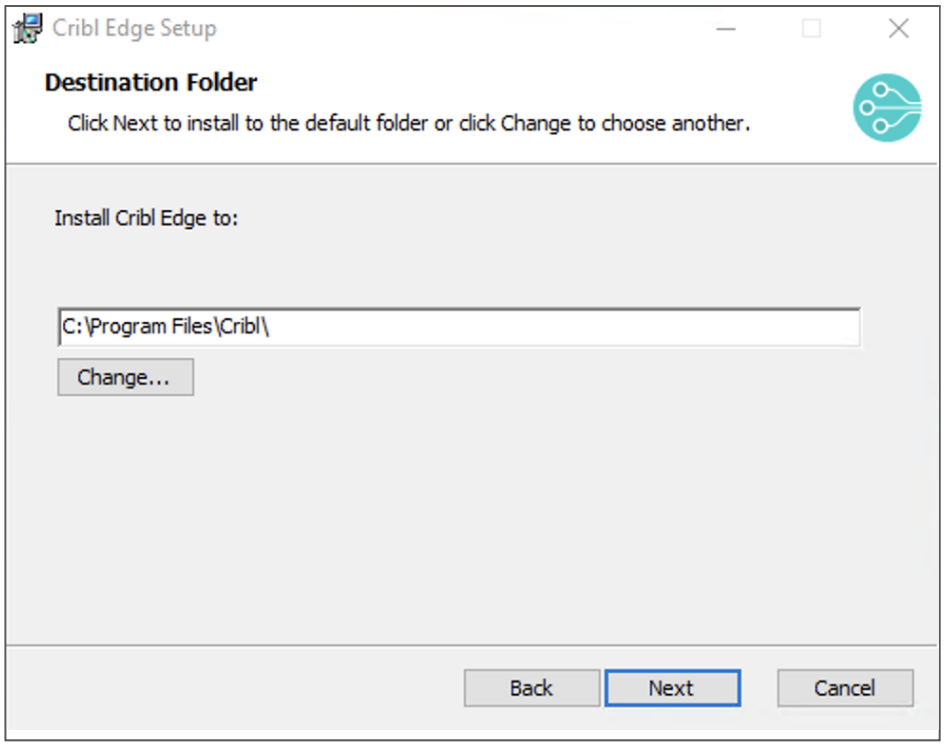
Either method installs Cribl Edge as a Windows service. This enables Cribl Edge to automatically restart whenever the Windows Server reboots.

If you are on Cribl.Cloud, the Enabling TLS After Installation section will walk you through creating an `instance.yml` file upon initial Cribl Edge/Windows installation, and then copying it to the same location for each subsequent install.

Please see Known Issues for any current limitations on automatic version upgrades.

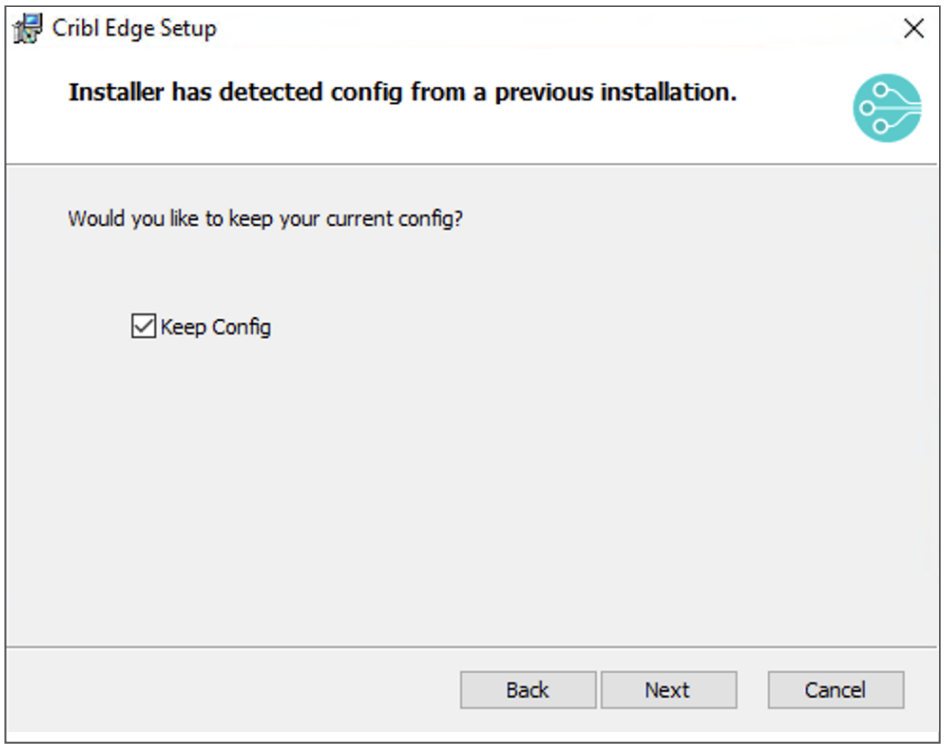
Using the Wizard

1. Double-click the Cribl Edge `.msi` to launch the Cribl Edge Setup Wizard. Click **Next** to start.
2. Read the **End-User License Agreement** and check the box to accept the terms. For the most current copy and details, see Terms of Service.
3. On the wizard's next page, confirm or change the **Destination Folder**. (The default path will expand as `C:\Program Files\Cribl\bin`).



Select location

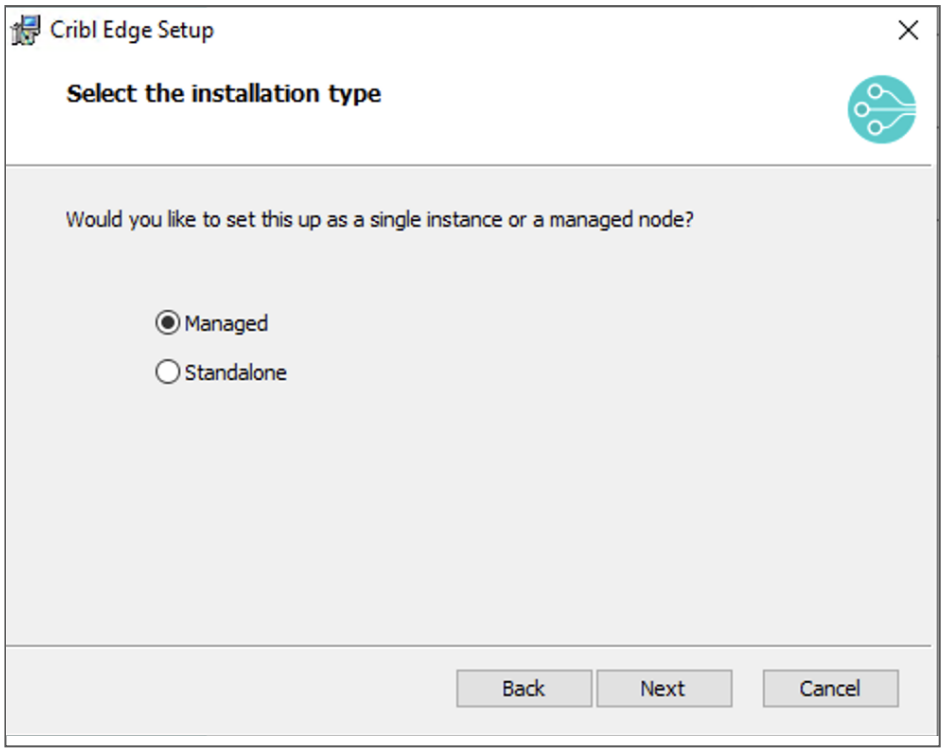
4. If the installer detects a config file from a previous installation (prior to v.4.1), the modal will ask you if you want to keep your current config. Check the box next to `Keep Config` to persist your previous configurations stored in the `instance.yml` file. Click **Next** to continue.



Keep Config

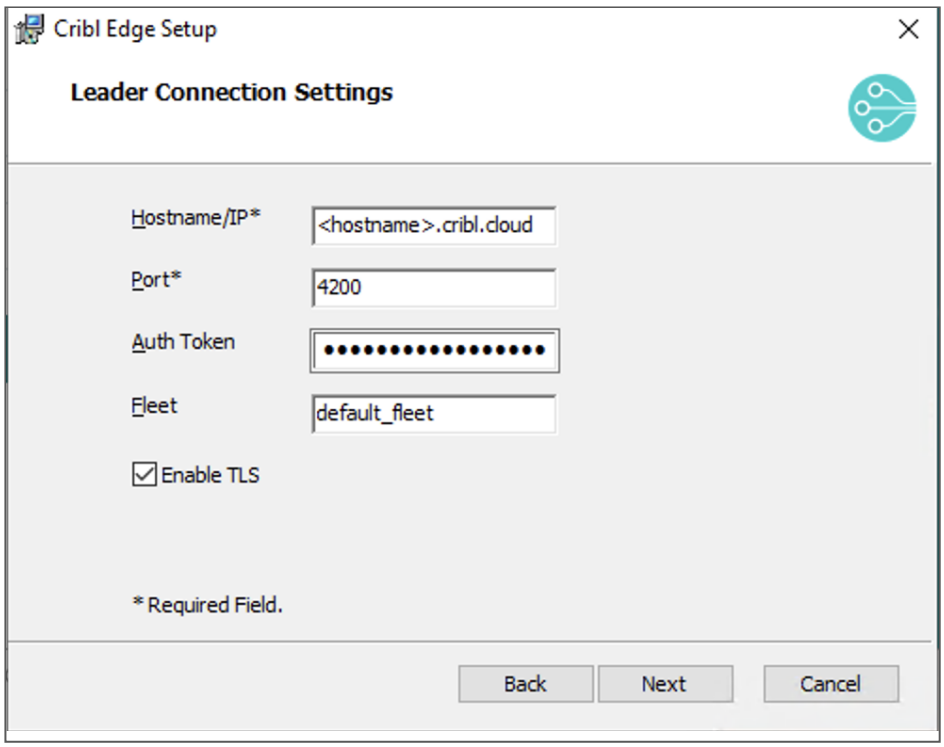
5. If you chose to keep your previous configuration, the installer will skip this option to **Select the installation type**. Select either **Managed** or **Standalone** as the installation type.

For Cribl.Cloud, Select **Managed Edge Node**.



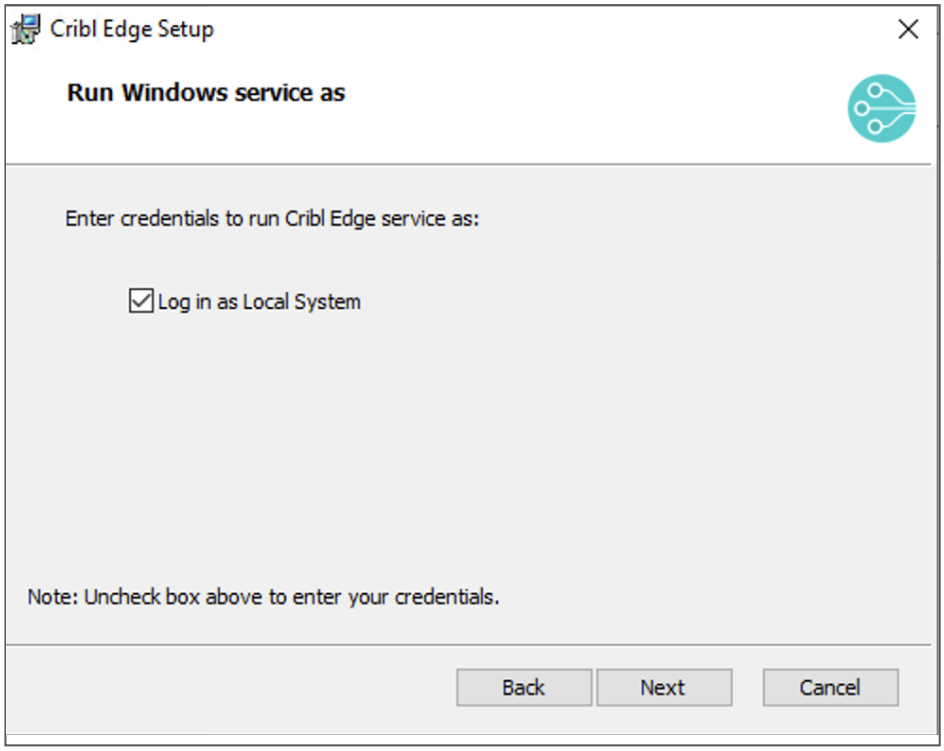
Select installation type

6. If you select **Managed** mode, enter the Leader URI (`<Leader-hostname-or-IP address>`), Auth Token, and optionally the Fleet.

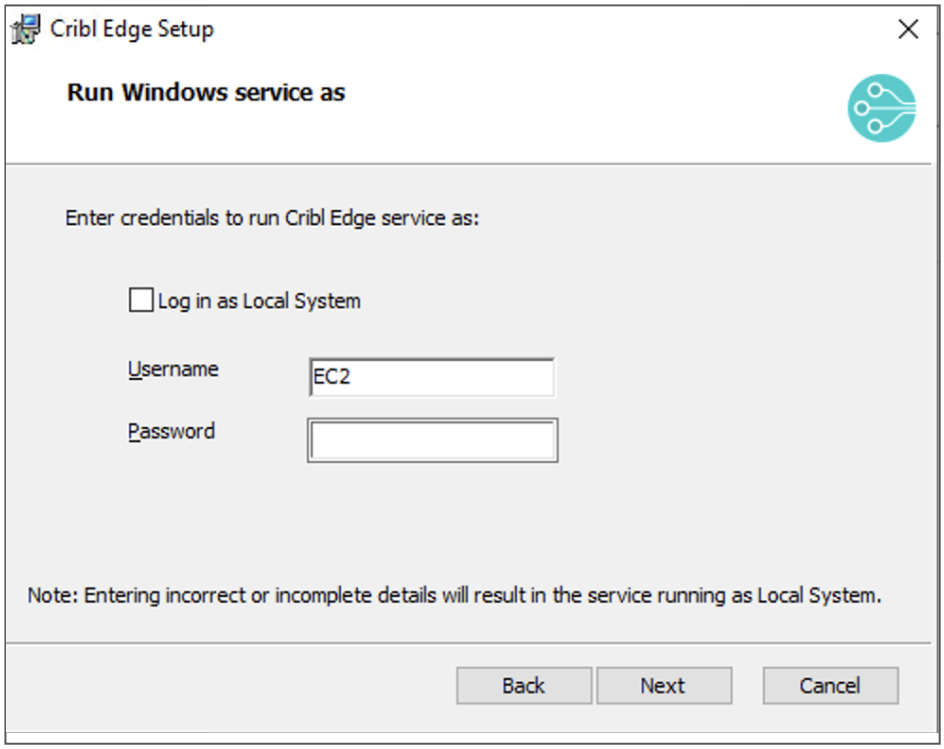


Connection Details

7. Optionally, **Log in as Local System** is checked by default. Uncheck it to enter your **Username** and **Password** credentials to run Cribl Edge. To successfully log Cribl Edge as a service, you might need to include the domain name in the **Username**. Click **Next** to continue.



Log in as Local System

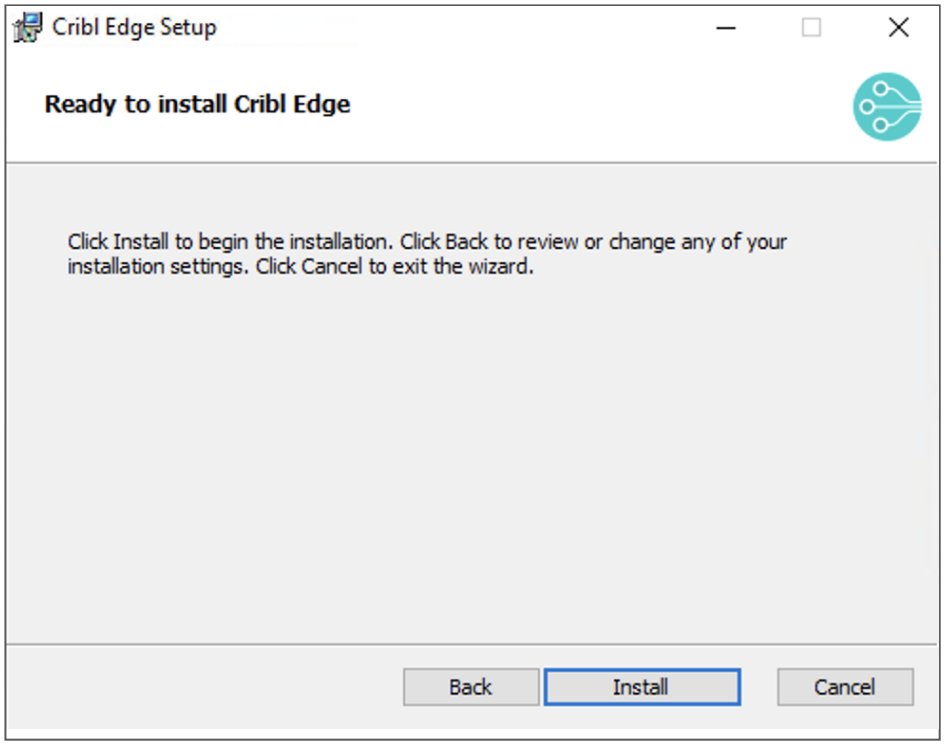


Use Credentails

8. Optionally select **Create a desktop shortcut**.

Desktop shortcut option

9. On the final **Ready to Install** page, click **Install** to confirm.



Installation summary

10. When installation is complete, double-click the Cribl Edge icon on your Desktop or File Explorer. This will launch Cribl Edge's login page in your default browser. Or go to `http://localhost:9420` and log in with default credentials (`admin:admin`).

As of Cribl Edge v.4.0, the modified data (including the `instance.yml`) now goes to `C:\ProgramData\Cribl` instead of `C:\Program Files\Cribl`.

Retrieving Cribl.Cloud Credentials

On Cribl.Cloud, retrieve the **Leader URI** and **Auth Token** from your Cloud instance like this:

1. Navigate to the Manage Edge Nodes page.
2. From the **Add/Update Edge Node** control at upper right, select **Bootstrap new**.
3. Select **Add Windows**.
4. Copy the **Leader hostname/IP** value.
5. The **Leader Port** field is `4200`.
6. Display the **Auth Token** value, and copy/paste it into the installer. The **Auth Token** is required to enable communication between the Leader and Edge Node.

Using the MSI Installer

From the command line, you can install Cribl Edge as a single instance or as a Managed Node.

Installing Single-Instance/Standalone Cribl Edge

To run Edge locally, as a single-instance deployment on your own machine, enter the following at your command prompt for a silent install:

```
msiexec /i cribl-<version>-<build>-<arch>.msi /qn
```

Next, go to `http://localhost:9420` and log in with default credentials (`admin:admin`).

You can now start configuring Cribl Edge with Sources and Destinations, or start creating Routes and Pipelines.

Installing Managed Edge Nodes

You can use the Leader UI to concatenate and copy/paste the bootstrap script for adding a Windows Node.

Or, to run Cribl Edge as a managed Node, enter the following at your command prompt for a silent install:

```
msiexec /i cribl-<version>-<build>.msi /qn MODE=mode-managed-edge HOSTNAME=<yourhos
```

For Cribl Edge v.4.1.0 or later, use the flag `KEEPDATA=1` to persist data between installs. `KEEPDATA` will not overwrite the existing configurations, state, logs, etc.

```
msiexec /i cribl-<version>-<build>.msi /qn KEEPDATA=1
```

For prior versions, use the flag `COPYDATA=1` to persist data between installs. `COPYDATA` copies the `instance.yml` file from `C:\Program Files\Cribl` to `C:\ProgramData\Cribl`.

You can now manage this node from the specified Leader.

For other parameter options, see the CLI Reference. Here is an example command:

```
msiexec /i cribl-<version>-<build>.msi /qn MODE=mode-managed-edge HOSTNAME=192.0.2.
```

If you are installing Cribl Edge into a Docker container on Windows, you must include the `username`. For example:

```
msiexec /i cribl-<version>-<build>.msi /qn MODE=mode-managed-edge HOSTNAME=192.0.2.
```

Enabling TLS After Installation

Some on-prem deployments don't require secure (TLS) communication between the Leader and Edge Node. In these cases, the Managed Node will complete its configuration, followed by removing the `admin/admin` password. There will be no reason to locally log in, as you can manage the Edge Node via the Leader UI.

For sites using secure TLS connections to the Leader, including Cribl.Cloud, you must configure the local Edge Node to enable TLS. To do this:

1. Log into your local instance (at `http://localhost:9420`) with the default credentials (`admin/admin`).
2. Enable TLS locally. For details, see Connecting to the Leader Securely.
3. When prompted, restart your Cribl Edge instance.
4. Locate the `instance.yml` file in:
    - Cribl Edge v.4.0.X or earlier: `Program Files\cribl\local\_system`.

- Cribl Edge v. 4.1 or later: `C:\ProgramData\Cribl`.

To avoid this post-installation update for subsequent Windows installations to other servers, you can copy the `instance.yml` from this server. After the installation is complete, paste the `instance.yml` file into subsequent servers' `\Program Files\cribl\local\_system` folder.

To connect each new Cribl Edge instance to this config file, you'll need to either restart the Windows Server, or simply run the following commands as a Windows Administrator:

```
net stop cribl

net start cribl
```

Troubleshooting and Logging

Here are some helpful tips:

Setting Logging Options

To debug `.msi` installation or upgrade issues, you can run `msiexec.exe` to set logging options. See Microsoft's Logging Options topic.

Changing the Installation Directory

If you want to change the installation directory, use the `msiexec` command `APPLICATIONROOTDIRECTORY`. For example:

```
msiexec.exe /i Cribl-Edge.msi APPLICATIONROOTDIRECTORY="C:\test\" /L*V "C:\Log\Crib
```

If you're using PowerShell to run the silent install, add two extra pairs of quotes to escape the command's own quote delimiters. For example:

```
msiexec /i cribl-<version>-<build>.msi """MODE=mode-managed-edge HOSTNAME=192.0.2.1
"""
```

Cribl University offers a course titled How to Install & Configure Edge on Windows that illustrates the installation steps. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions**

presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

Frequently Asked Questions

Here are some common questions about Cribl Edge on Windows, and the answers!

Q: What is the `nssm.exe` process?

A: `nssm.exe` is a wrapper for the `cribl.exe` binary. It gives Windows Service Manager the ability to manage an executable program without having to provide a number of required binary APIs on the executable. So, instead of the `cribl.exe` binary needing to provide the APIs, `nssm.exe` acts as an intermediary, providing the necessary APIs itself and subsequently spawning `cribl.exe` as a child process.

Q: Why is `nssm.exe` important?

A: The `nssm.exe` process ensures that Cribl Edge is stable and reliable on Windows, both by booting up Cribl Edge and restarting it for you. With the `nssm.exe` process running, Cribl Edge remains operational even if it closes unexpectedly. While some antivirus software might flag `nssm.exe` due to its monitoring and restart functionality, `nssm.exe` is a legitimate and essential component of Cribl Edge.

High CPU utilization may appear shortly after launching Cribl Edge but is not necessarily related to `nssm.exe`. We recommend waiting for Cribl Edge to fully boot and stabilize before checking CPU usage.

3.3.1. System Proxy Configuration On Windows

You can direct all outbound HTTP/S requests to go through proxy servers.

Enable Proxy Server

First, you need to make sure the proxy server is enabled.

There are a few ways to do it, including using Windows system settings, Group Policy Objects, and environment variables.

Windows Proxy Settings

To enable the proxy server for all processes and services on a host, you can use the system proxy settings:

1. In Windows settings go to **Network & Internet** > **Proxy**.
2. Under **Manual proxy setup**, toggle **Use a proxy server** to **On**.
3. In the **Address** and **Port** boxes, enter the proxy server name or IP address and (optionally) port.
4. Confirm with **Save**.

Group Policy Object

Alternatively, you can use a Windows Group Policy Object to enable the proxy server across groups of accounts or groups of servers. See Microsoft Q&A thread for a detailed instruction.

Environment Variables

Another way to enable proxy is to set the `HTTP_PROXY` and `HTTPS_PROXY` environment variables on the Cribl Edge Windows Service before starting Cribl Edge.

Configure the variables, either with your proxy's IP address, or with a DNS name that resolves to that IP address. Optionally, follow either convention with a colon and the port number to which you want to send queries.

```
$environment = [string[]]@("HTTP_PROXY=http://proxy:1234,HTTPS_PROXY=http://proxy:1
Set-ItemProperty HKLM:SYSTEM\CurrentControlSet\Services\Cribl -Name Environment -Va
```

In the above example, note that when you set an `HTTPS_PROXY` environment variable, the referenced URL should generally be in `http` format.

Restarts and Case Conflicts

You must restart Cribl Edge on the affected Nodes if the application is running when you first configure variables or make changes to them.

The environment variables' names can be either uppercase or lowercase. However, if you set duplicate versions of the same name, the lowercase version takes precedence. E.g., if you've set both `HTTPS_PROXY` and `https_proxy`, the IP address specified in `https_proxy` will take effect.

Communication with Leader over Proxy

If you outbound TCP connection from an Edge Node to the Leader must go through proxy, you need to set the `CRIBL_DIST_WORKER_PROXY` environment variable:

```
CRIBL_DIST_WORKER_PROXY=<socks4|socks5>://<username>:<password>@<host>:<port>
```

HTTP and/or HTTPS?

Several Cribl Edge endpoints rely on the HTTPS protocol – the Cribl telemetry endpoint – which must be accessed with some license types, as well as the CDN used to propagate application updates, certain documentation features (API Reference and docs PDFs), and config bundle downloads.

You might configure certain other Cribl Edge features that require access to HTTP endpoints. For maximum flexibility, consider setting environment variables to handle both the HTTPS and HTTP protocols.

Authenticating on Proxies

You can use HTTP Basic authentication on HTTP or HTTPS proxies. Specify the user name and password in the proxy URL. For example:

```
$environment = [string[]]@("HTTP_PROXY=http://username:password@proxy:1234,HTTPS_PR
Set-ItemProperty HKLM:SYSTEM\CurrentControlSet\Services\Cribl -Name Environment -Va
```

Bypassing Proxies with `NO_PROXY`

If you've set the above environment variables, you can negate them for specified (or all) hosts. Add the `NO_PROXY` environment variable to identify URLs that should bypass the proxy server, to instead be sent as direct requests. Use the following format:

```
NO_PROXY="<list of hosts/domains>"
```

Cribl recommends including the Leader Node's host name in the `NO_PROXY` list.

`NO_PROXY` Usage

Within the `NO_PROXY` list, separate the host/domain names with commas or spaces.

Optionally, you can follow each host/domain entry with a port. If not specified, the protocol's default port is assumed.

To match subdomains, you must either list them all in full (for example, `NO_PROXY=foo.example.com,bar.example.com`), or apply a wildcard by prefixing the domain name with a period or `*.`: `NO-PROXY=".example.com` or `NO-PROXY="*.example.com`.

To match the whole domain including its subdomains, add it both with and without wildcard to the list: `NO-PROXY="example.com,.example.com`.

To disable all proxies, use the `*` wildcard: `NO_PROXY="*"`. `NO_PROXY` with an empty list disables no proxies.

Cloud `NO_PROXY` Usage

You must include any cloud metadata endpoints (such as the AWS Instance Metadata Service) in the `NO_PROXY` list:

- AWS EC2 and Azure VM instances must include `169.254.169.254` in the list. If using IPv6 on AWS EC2, add `fd00:ec2::254` to the list.
- AWS ECS Fargate tasks must include `169.254.170.2`.
- GCP (Google Cloud Platform) VM instances must include `metadata.google.internal` and `169.254.169.254`.

Where Proxies Apply

Proxy configuration is relevant to the following Cribl Edge components that make outbound HTTP/S requests:

Destinations

- S3 Compatible Stores
- AWS Kinesis Streams
- AWS CloudWatch Logs
- AWS SQS
- Azure Blob Storage
- Azure Monitor Logs
- Cribl HTTP
- CrowdStrike Falcon LogScale
- Elasticsearch
- Grafana Cloud
- Honeycomb
- Loki
- Prometheus
- Splunk HEC
- Webhook

Notification Targets

- PagerDuty
- Webhook

Testing Proxies

To initially test your proxy configuration, consider setting up a simple, free proxy server like mitmproxy (https://mitmproxy.org/), and then monitoring traffic through that server. Verify that you can trace proxied requests from your Cribl Edge instance, and can validate that outgoing requests (to Destinations) are working properly.

3.4. Running In A Docker Container

As a best practice, we recommend using the UI to concatenate and copy/paste the bootstrap script for adding a Docker node.

Or, for a single-instance deployment, start the container with the Docker command below:

```
docker run -d -e CRIBL_EDGE=1 \
 -p 9420:9420 \
 -v /var/run/appscope:/var/run/appscope \
 -v /var/run/docker.sock:/var/run/docker.sock \
 -v /:/hostfs:ro \
 --privileged \
 --restart unless-stopped \
 --name cribl-edge \
 cribl/cribl:latest
```

For a distributed deployment, edit the Docker command below to replace `INSERT_TOKEN` with a unique, secure token value. Then start the container with the revised command.

```
docker run -d \
 --privileged \
 -e CRIBL_DIST_MODE=managed-edge \
 -e CRIBL_DIST_MASTER_URL=tcp://INSERT_TOKEN@leaderhere:4200?group=fleet_name
 -e "CRIBL_EDGE=1"
 -v /var/run/appscope:/var/run/appscope \
 -v /var/run/docker.sock:/var/run/docker.sock \
 -v /:/hostfs:ro \
 -p 9420:9420
 --restart unless-stopped \
 --name cribl-edge \
 cribl/cribl:latest
```

With a Leader on Cribl.Cloud, encryption is enabled by default. Set the hybrid worker's `CRIBL_DIST_MASTER_URL` environment variable to begin with the `tls://` protocol. For example:

```
CRIBL_DIST_MASTER_URL=tls://<token>@logstream-<tenant>.cribl.cloud:4200
```

Once that's running, the UI displays on `http://localhost:9420/`

The `-v /var/run/appscope:/var/run/appscope \` line exposes the Unix domain socket for the AppScope Source so that it's accessible to the `scope` application on the host, and in other containers that mount that same directory. See Running AppScope With Cribl Edge in a Docker Container in the AppScope documentation.

You can now start configuring Cribl Edge with Sources and Destinations, or start creating Routes and Pipelines.

Overriding Default Ports

Default ports can be overridden in the following configuration files:

- Edge UI port (`9420`): Default definitions for `host`, `port`, and other settings are set in `$CRIBL_HOME/default/cribl/cribl.yml`, and can be overridden by defining alternatives in `$CRIBL_HOME/local/cribl/cribl.yml`.
- Data Ports, for example HTTP In (`10080`), TCPJSON in (`10420`) have default definitions for `host`, `port` and other settings defined in `$CRIBL_HOME/default/cribl/inputs.yml`, and can be overridden by defining alternatives in `$CRIBL_HOME/local/cribl/inputs.yml`.

In the case of an API port conflict, the process will retry binding for 10 minutes before exiting.

Updating the Docker Image

Cribl recommends that you always use our latest stable container image wherever possible. This will provide bug fixes and security patches for any vulnerabilities that Cribl has discovered when scanning the base image OS, dependencies, and our own software.

You can explicitly pull the latest stable image with this CLI command:

```
docker pull cribl/cribl:latest
```

Updating the Packaged OS

Cribl strongly recommends that you monitor and patch vulnerabilities in the packaged OS. The base OS might have been updated with fixes for new vulnerabilities discovered after Cribl published its container images. This is especially important if you choose to keep an earlier Cribl image in production.

You can update the base OS image by updating the package, as shown in the following `Dockerfile`. This example assumes you're updating Cribl's `latest` image:

```
FROM cribl/cribl:latest
RUN apt-get update && \
    apt-get -y upgrade dpkg
```

Cribl Edge supports both Docker and `containerd` runtimes.

Troubleshooting

When you deploy containers via Kubernetes in `managed-edge` mode, this Node will – after applying the first config bundle from the Leader – listen on the Fleet's configured **API Server Settings** > **Host**.

If you encounter spurious health-check failures, change that **Host** setting to `0.0.0.0` (rather than the default `127.0.0.1`, which can block health checks).

Cribl University offers a course titled How to Install & Configure Edge Using Docker that illustrates the installation steps. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

3.5. Deploying Via Kubernetes

This page outlines how to deploy Cribl Edge in Kubernetes Pods to collect logs and metrics.

Cribl Edge supports both Docker and `containerd` runtimes.

Considerations

Cribl Edge connects to the Kubernetes Cluster API and the `kubelet` API on each node to read logs and metrics. To prevent under- or over-collection of logs and metrics, you must deploy the Cribl Edge agent as a DaemonSet in your Kubernetes cluster. Ensure that Cribl Edge is able to detect that it's running inside a Kubernetes Pod, by setting the `CRIBL_K8S_POD` environment variable.

To run the Kubernetes Sources (Metrics, Logs) inside a cluster, use the default `ServiceAccount`, and mount the connection details in `var/run/secrets/kubernetes.io/serviceaccount`. If you want the Source to look in a different directory, use the `CRIBL_SERVICEACCOUNT_PATH` environment variable to set the new path to the `ServiceAccount`.

Optionally, you can run the Kubernetes Sources (Metrics, Logs) outside of a cluster, by specifying connection details in a `kubeconfig` file (like `kubectl`). Either place the file in `$HOME/.kube/config`, or use the `$KUBECONFIG` variable to point to it.

If you are enabling role-based access control (RBAC) in your Kubernetes cluster, Cribl Edge requires a `ClusterRole` to authorize the Service Account to read logs and metrics. This is a `read-only` role, which does not require `write` permissions. Plan ahead with your Kubernetes admins and security team to authorize the installation of this `ClusterRole`.

If your Kubernetes deployment doesn't use valid certificates, then set the `CRIBL_K8S_TLS_REJECT_UNAUTHORIZED` environment variable to `0` to disable that expectation. When you disable this environment variable, all Kubernetes features (including Metadata, Metrics, Logs, and AppScope metadata) will tolerate invalid TLS certificates (i.e., expired, self-signed, etc.) when connecting to the Kubernetes APIs. If this environment variable is not defined or set to `1` and the certificate validation fails, then all connection attempts to the Kubernetes API will be rejected. For details, see Which Certificate Does kubelet Use?

All other Cribl Edge Kubernetes components can run inside a namespace of your choice. This page uses the example namespace `cribl-edge`.

Deploying

To deploy the Cribl Edge agent in your Kubernetes cluster, you can use our Helm charts. These charts provide the minimum configuration required to install Cribl Edge correctly.

You can customize our Helm charts with additional configurations required for your Kubernetes platform – such as annotations for load balancers, environment variables for secrets, etc. For details, follow the link above.

A simple deployment example using Cribl's Helm chart:

```
helm repo add cribl https://criblio.github.io/helm-charts/

helm install -n cribl-edge --set "cribl.leader=tcp://<token>@leader:4200" cribl-edg
```

You can also use the UI to concatenate and copy/paste the bootstrap script for adding a Kubernetes node.

If you are building your own manifests, the DaemonSet pods must be configured with the following environment variables:

- `CRIBL_DIST_LEADER_URL: <changeme>`
- `CRIBL_DIST_MODE: managed-edge`
- `CRIBL_EDGE: true`
- `CRIBL_K8S_POD :<must match the name of the pod>`

Tolerations

By default, Kubernetes will not schedule a Pod on a node that has taints. You must configure your deployment to allow scheduling on these nodes; otherwise, Cribl Edge will be unable to collect logs and metrics from them.

The defaults in our Helm chart's `values.yaml` file are configured like this:

```
tolerations:
  - operator: Exists
```

RBAC

Cribl Edge's Kubernetes Logs and Kubernetes Metrics Sources use the Kubernetes API to read logs and metrics from the cluster. If you have role-based access control (RBAC) implemented, you will need to authorize the Cribl Edge Pod Service Account to read the relevant data.

Kubernetes Events Source

To collect cluster-level events, the Kubernetes Events Source needs watch access to the Kubernetes API.

An example RBAC rule is:

```
- apiGroups:
    - "events.k8s.io"
  resources:
    - events
  verbs: ['watch']
```

Kubernetes Logs Source

The Kubernetes Logs Source requires the ability to get, list, and watch Pods in all namespaces.

An example RBAC rule is:

```
- apiGroups:
    - ""
  resources:
    - pods
  verbs: ['get', 'list', 'watch']
```

Kubernetes Metrics Source

The Kubernetes Metrics Source reads metric information from many resources in your cluster. The minimum permissions required are:

```
- apiGroups:
    - ""
  resources:
    - configmaps
    - endpoints
    - limitranges
    - namespaces
    - nodes
    - persistentvolumeclaims
    - pods
    - replicationcontrollers
    - secrets
    - services
    - strategicMergePatches
    - nodes/log
    - nodes/metrics
    - nodes/proxy
    - nodes/spec
    - nodes/stats
  verbs: ['get', 'list', 'watch']

- apiGroups:
    - "apps"
  resources:
    - daemonsets
    - deployments
    - replicasets
    - statefulsets
  verbs: ['get', 'list', 'watch']
- apiGroups:
    - "batch"
  resources:
    - cronjobs
    - jobs
  verbs: ['get', 'list', 'watch']
- apiGroups:
    - "autoscaling"
  resources:
    - horizontalpodautoscalers
  verbs: ['get', 'list', 'watch']
- apiGroups:
    - "policy"
```

```
    resources:
      - poddisruptionbudgets
    verbs: ['get', 'list', 'watch']
  - apiGroups:
      - "networking.k8s.io"
    resources:
      - ingresses
      - networkpolicies
    verbs: ['get', 'list', 'watch']
  - apiGroups:
      - "admissionregistration.k8s.io"
    resources:
      - mutatingwebhookconfigurations
      - validatingwebhookconfigurations
    verbs: ['get', 'list', 'watch']
  - apiGroups:
      - "certificates.k8s.io"
    resources:
      - certificatesigningrequests
    verbs: ['get', 'list', 'watch']
  - apiGroups:
      - "storage.k8s.io"
    resources:
      - storageclasses
      - volumeattachments
    verbs: ['get', 'list', 'watch']
```

Prometheus Edge Scraper

To discover a Kubernetes Node or Pods, the Prometheus Edge Scraper Source must be able to list Pods across all namespaces.

An example RBAC rule is:

```
  - apiGroups:
      - ""
    resources:
      - pods
    verbs: ['list']
```

Health Checks

By default, Cribl Edge's API listens on port `9420`, instead of on Cribl Stream's default `9000` port. This requires special care when you configure your Edge Fleet in the Cribl UI.

By default, the API Server is configured to listen on the loopback interface (`127.0.0.1`). Change the listening interface address to `0.0.0.0` in **Fleet Settings** > **System** > **General Settings** > **API Server Settings** > **General**. This will bind to all IPs in the container. If you do not make this change, your health checks will start failing after the Edge agent downloads its initial config bundle.

An example health-check configuration for Kubernetes is:

```
readinessProbe:
    httpGet:
        path: /api/v1/health
        port: 9420
        scheme: HTTP
    initialDelaySeconds: 15
    timeoutSeconds: 1
livenessProbe:
    httpGet:
        path: /api/v1/health
        port: 9420
        scheme: HTTP
    initialDelaySeconds: 15
    timeoutSeconds: 1
```

Environment Variables

You can use the following environment variables to configure Cribl Edge on Kubernetes:

- `CRIBL_SERVICEACCOUNT_PATH` - by default, set to: `/var/run/secrets/kubernetes.io/serviceaccount`.
- `CRIBL_K8S_POD` - the name of the Kubernetes Pod in which Cribl Edge is deployed.
- `KUBECONFIG` - set this for local testing.
- `CRIBL_K8S_FOOTGUN`- set to `true` to enable resource-intensive, potentially risky modes of the Kubernetes Metrics and Kubernetes Logs Sources.
- `CRIBL_K8S_TLS_REJECT_UNAUTHORIZED` - set to `0` to disable certification validation when connecting to the Kubernetes APIs. When you disable this environment variable, all Kubernetes features (including Metadata, Metrics, Logs, and AppScope metadata) will tolerate invalid TLS certificates (i.e., expired, self-signed, etc.) when connecting to the Kubernetes APIs.

See also our list of all Cribl Environment Variables.

Uninstalling

To remove the Cribl Edge agent from your Kubernetes cluster, run the following command:

```
helm uninstall -n <namespace> <releasename>
```

The helm uninstall command automatically deletes all deployed resources associated with the Cribl Edge Helm Chart.

Cribl University offers two courses that provide a good overview of working with Kubernetes: How to Install & Configure Edge on Kubernetes and Collecting Data from K8s. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

4. Fleet Management

Getting started with managing Fleets of Edge Nodes in a distributed deployment

---

In a distributed environment, Edge Nodes can be managed centrally by a single Leader Node, which is responsible for keeping configurations in sync, tracking edge node status and monitoring metrics.

Concepts

Single Edge Node – a single Cribl Edge instance, running as a standalone (not distributed) installation on one server/endpoint.

Leader Node – an instance running in **Leader** mode, used to centrally author configurations and monitor Edge Nodes in a distributed deployment.

Managed Edge Node – an instance running as a **Managed Edge**, whose configuration is fully managed by a Leader Node. (By default, will poll the Leader for configuration changes every 10 seconds.)

Fleet – a collection of Edge Nodes that share the same configuration. You map Nodes to a Fleet using a Mapping Ruleset.

Subfleet – a Subfleet groups and manages Edge Nodes that share the same configuration. Each Subfleet inherits configurations from its parent Fleet. Updating and deploying parent-level configurations applies the changes to the Subfleets. The Subfleets then deploy the configuration changes to their Edge Nodes.

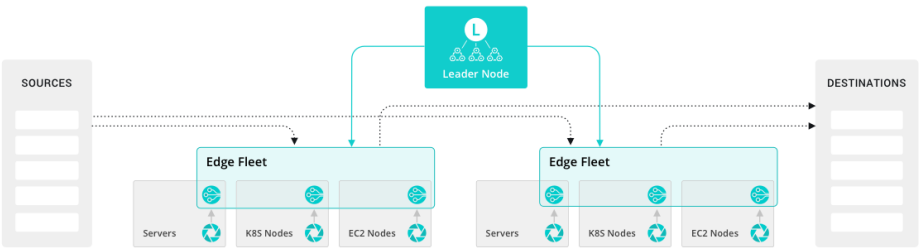
Worker Process – a Linux process within a Single Instance, or within Edge Nodes, that handles data inputs, processing, and output. The process count for Edge Nodes is constrained to 1.

Mapping Ruleset – an ordered list of filters, used to map Nodes into Fleets.

Multiple Fleets are very useful in making your configuration reflect organizational or geographic constraints. E.g., you might have a U.S. Fleet with certain TLS certificates and output settings, versus an APAC Fleet and an EMEA Fleet, each with their own distinct certs and settings.

Architecture

This is an overview of a distributed Edge deployment's basic components. (The type/s of Nodes within each Fleet will vary depending on your overall architecture:)



Distributed deployment architecture

Leader Node

- API Process – Handles all the API interactions.
- *N* Config Helpers – One process per Fleet. Helps with maintaining configs, previews, etc.

Edge Node

- Single Process – Handles everything; collection, processing and communication with the Leader Node.

Leader Node Requirements

- **OS**: Linux: RedHat, CentOS, Ubuntu, AWS Linux (64bit)
- **System**: +4 physical cores, +8GB RAM, 5GB free disk space
- **Network**:
  - Heartbeat Port: `4200`: managed Edge nodes communicate with the Leader Node on port `4200` by default.
  - UI/API Port: `9000`: users acccess the Leader on port `9000` by default.
- **Git**: `git` must be available on the Leader Node. See details below.
- **Browser Support**: The five most-recent versions of Chrome, Firefox, Safari, and Microsoft Edge.

We assume that 1 physical core is equivalent to 2 virtual/hyperthreaded CPUs (vCPUs). All quantities listed above are minimum requirements. We recommend deploying the Leader on stable, highly available infrastructure, because of its role in coordinating all Edge instances.

Edge Node Requirements

See Deployment Planning for requirements and other details.

Installing on Linux

See Installing on Linux.

Installing on Windows

See Installing on Windows.

Version Control with `git`

Leader Node requires `git` (version 1.8.3.1 or higher) to be available locally on the host.

Configuration changes must be committed to git before they're deployed.

If you don't have `git` installed, check here for details on how to get started.

The Leader node uses `git` to:

- Manage configuration versions across Fleets.
- Provide users with an audit trail of all configuration changes.
- Allow users to display diffs between current and previous config versions.

4.1. Fleets

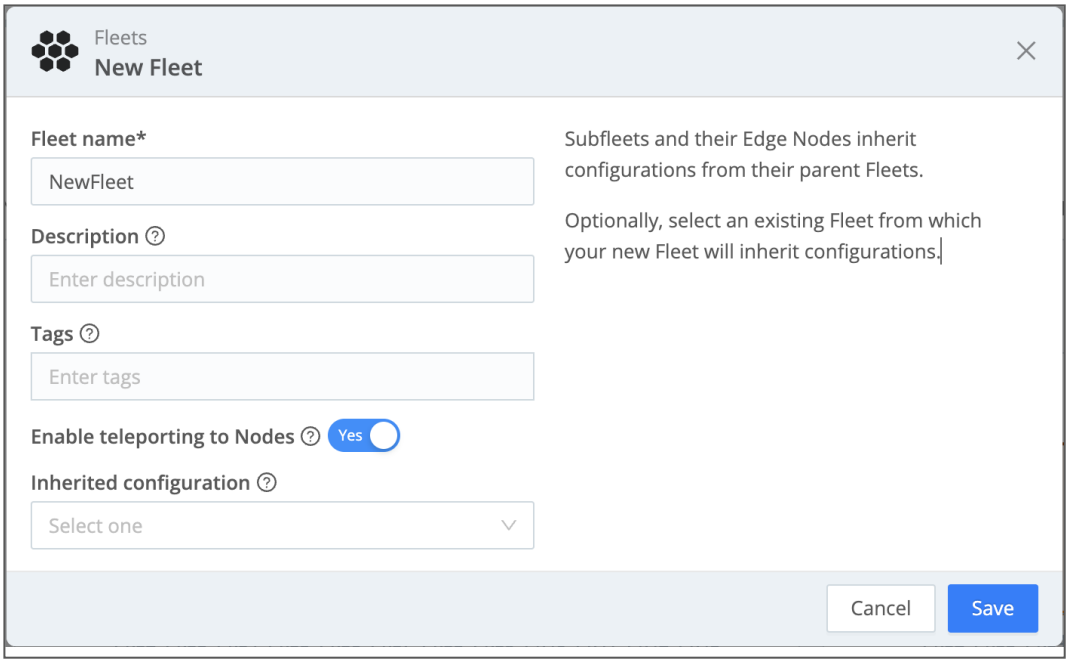
Fleets and their corresponding Subfleets allow you to author and manage configuration settings for a particular set of Edge Nodes, so you can flexibly group Edge Nodes into logical Fleets and Subfleets. These Fleets and Subfleets can share and reuse configurations.

With Cribl Edge v.4.2.x and later, you can assign Fleet-level permissions to users. For details, see Members and Permissions.

Creating a Fleet

To create a new Fleet:

1. Click the **Manage** tab in the top nav.
2. In the **Fleets** lower tab, click **New Fleet**.
3. Enter a descriptive **Fleet name**, and optionally, a relevant **Description** and Tags.
   To enable teleporting to the Edge Nodes via the Leader's UI, toggle **Enable teleporting to Nodes** to **Yes**, then click **Save**.



Add a new Fleet

Subfleets and Inheritance Configurations

You can organize your Fleets and Subfleets into a hierarchy of configuration layers from the top level down. At the Fleet level, this might include grouping Edge Nodes with basic configurations like common logging

locations, metrics, as well as common Sources and Destinations. At the Subfleet level, you can group Edge Nodes to pick up configurations specific to the applications and services that are running on the Nodes.

There are several ways you can build the hierarchy:

- Organizationally.
- Geographically (if your organization collects different data categories in different regions).
- Data center–based.
- OS-based.

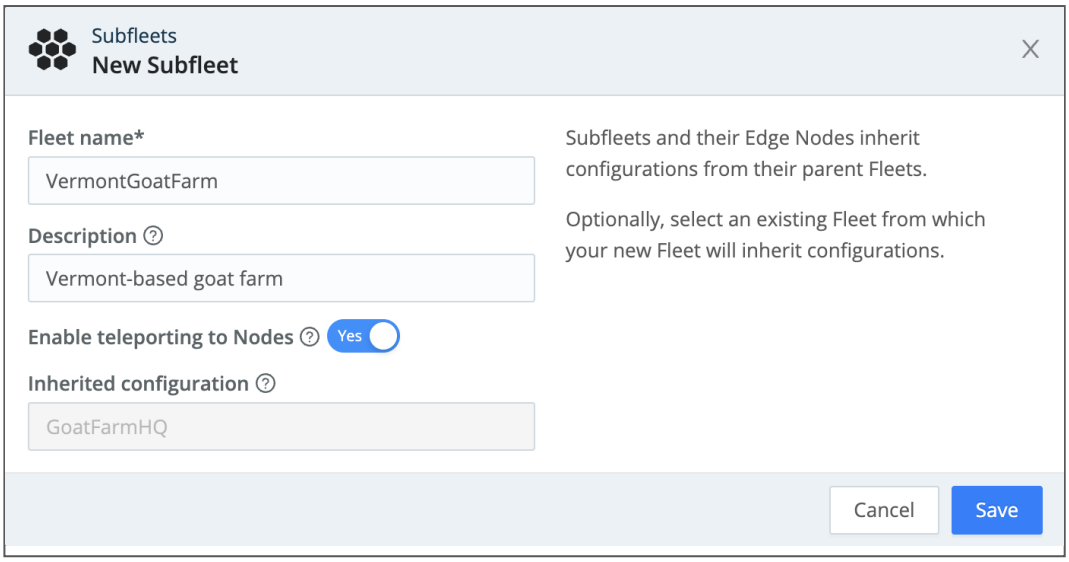
Layering the configurations allows you to manage a large number of Edge Nodes that share common yet differentiated configurations. For example, as an administrator, you might need multiple layers of configurations to manage the following `Org > Department > OS > Application`. In this case, the Edge Nodes at the `Application` level can inherit and share configurations from the `Org` level. You can update a parent-level configuration (Fleet) and have it applied to all Subfleets and their respective Edge Nodes, reducing the time and effort needed to manage them. For guidance, see our better practices doc: Fleets Hierarchy and Design.

Adding a Subfleet

To add a Subfleet:

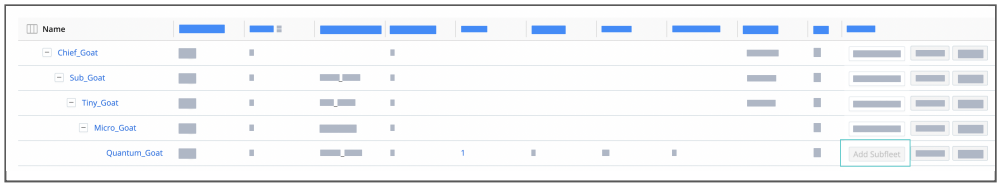
1. In the row of your desired Fleet, click **Add Subfleet** in the **Actions** column.
2. Enter a descriptive **Fleet name** and optional **Description** and **Tags**.
3. Select an existing Fleet from the **Inherited configuration** drop-down. The Subfleet will inherit its configuration from the selected Fleet.
4. Click **Save**.

You can also add a new Subfleet from a parent Fleet by clicking the **Add Subfleet** button in the **Actions** column.



Add a Subfleet

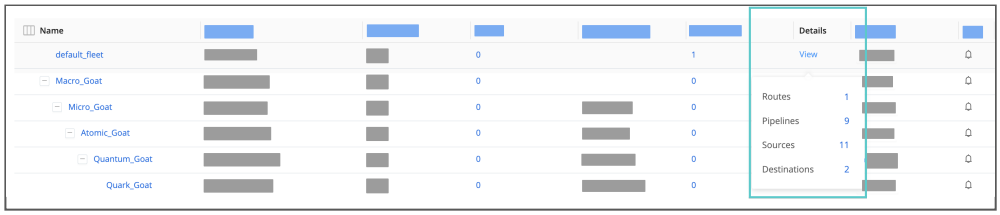
Similarly, you can create a child Fleet in a selected Subfleet. Subfleets can inherit configurations to five levels, at which point the **Add Subfleet** button will be disabled.



Subfleet's maximum depth

Viewing Subfleet Details and Options

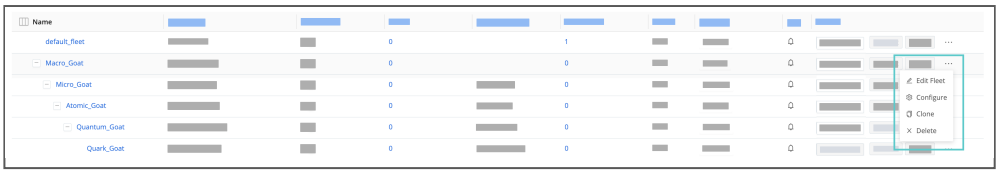
To display the Fleet's number of Routes, Pipelines, Sources, and Destinations, click **View** in a Fleet's or Subfleet's **Details** column.



View Subfleet details

On Cribl.Cloud, you'll also see a **Group Type** column, showing whether the deployment is **Hybrid** or **Cloud**.

To display the **Edit Fleet**, Configure, **Clone**, and **Delete** options, click the Fleet's or Subfleet's ••• (Options) menu.

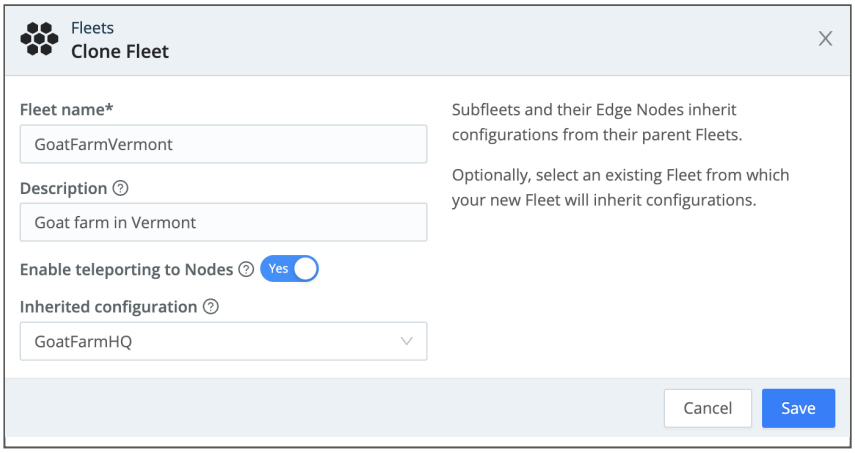


Edit, Configure, Clone, and Delete Fleets and Subfleets

Cloning and Deleting a Subfleet

Cloning a Fleet copies the Fleet's settings, while cloning a Subfleet inherits the parent's configurations.

To clone a Subfleet, click its **Options (•••)** menu and select **Clone**.



Clone a Subfleet

To delete a Fleet or Subfleet, click **Delete** in the **Options (•••)** menu. You will be prompted to confirm your decision to delete the Fleet/Subfleet, click **Yes** to confirm.

You can also commit and deploy changes to Fleets and Subfleets on this page. For details, see Committing and Deploying Fleet Configurations.

Configuring multiple Fleets requires an Edge Enterprise or Standard license.

Accessing Inherited Packs, Lookups, and Samples

Fleets with inherited configurations can also inherit Packs, Lookups, and Samples. Here's how to access them:

- Access inherited Packs in **More** > **Data Routes**, then open the **Pipeline** drop-down in a Route. Inherited Packs don't appear in the Packs list.

- You can teleport into a Node and view Lookups and Samples in their respective lists when you deploy configuration changes to the Edge Nodes.

Cribl University offers two courses that provide a good overview of Fleets: Fleet Administration & Management and Fleet Hierarchy. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

Configuring a Fleet

Navigate to a Fleet/Subfleet, click the **Options (•••)** menu to display the **Configure** option. Click **Configure** to display an interface for authoring and validating its configuration. You can configure everything for this Fleet as if it were a single Edge instance – using a similar visual interface.

Can't Log into the Edge Node as Admin User?

To explicitly set passwords for Fleets, see User Authentication.

Committing and Deploying Fleet Configurations

To make new configurations visible and available to Subfleets, you must first **Commit** the changes within parent Fleets' configurations.

On Cribl Edge v.4.1.x and later, when you commit changes to a parent Fleet, you now have the option to **Commit and Deploy with Subfleets**. When you **Deploy with Subfleets**, the Subfleets will receive the new inherited configuration. For guidelines on deploying to a large number of Edge Nodes, see Leader Scalability.

Deploy with Subfleets

Recovering a Fleet

If you delete a Fleet, you lose the context required to access the Version Control menu. As a result, you also lose the option to recover the Fleet before it was deleted.

You can recover a deleted Fleet using one of the following two methods:

- Creating a New Fleet with the Same Name
- Temporarily Stopping the Cribl Server

Creating a New Fleet with the Same Name

1. Create a new Fleet with the same name as the one that was deleted. This action will restore the deleted Fleet to the system.
2. Select a previous version of the deleted Fleet from the Version Control menu if you previously committed the deletion of your Fleet.

If you did not previously commit the deletion of your Fleet, the system will automatically restore it.

Temporarily Stopping the Cribl Server

1. Stop the Cribl server.
2. Identify the last commit before the Fleet was deleted.
3. Revert to that commit.
4. Restart the Cribl server.

Mapping Edge Nodes to Fleets

Mapping Rulesets are used to map Edge Nodes to Fleets. Within a ruleset, a list of rules evaluate Filter expressions on the information that Edge Nodes send to the Leader.

Only one Mapping Ruleset can be active at any one time, although a ruleset can contain multiple rules. At least one Fleet should be defined and present in the system.

In a ruleset, the order of Rules matters. The **Filter** section supports full JS expressions. The ruleset matching strategy is first-match, and one Edge Node can belong to only one Fleet.

Managing Edge Nodes on Multiple Platforms

The Leader is unaware of Edge Nodes' platforms (i.e, Linux or Windows) within a Fleet, which means the ConfigHelper omits platform-specific limitations. Therefore, when you manage Edge Nodes on heterogeneous platforms, create a Windows-specific Fleet and mapping.

For mapping details, see the Mapping sections below. For what's supported on Windows, see Cribl Edge on Windows.

Creating a Mapping Ruleset

To create a Mapping Ruleset:

1. Click **Manage** on the top nav > **Mappings**.
2. Click **Add Ruleset**.
3. Give the **New Ruleset** a unique **ID** and click **Save**.
4. Click the **Configure** button and start adding rules with **Add Rule**.

While you build and refine rules, the Preview in the right pane will show which currently reporting and tracked workers map to which Fleets.

A ruleset must be activated before it can be used by the Leader. To activate it, go to **Mappings** and click **Activate** on the required ruleset. The **Activate** button will then change to an **Active** toggle. Using the adjacent buttons, you can also **Configure** or **Delete** a ruleset, or **Clone** a ruleset if you'd like to work on it offline, test different filters, etc.

Although not required, Edge Nodes can be configured to send a preferred Fleet name with their payload (found in payload's `cribl.group` key) See below how this ranks in mapping priority.

For platform-specific mapping, set the `platform` property to `win32` for Windows, or to `linux` for Linux.

Add a Mapping Rule – Example

Within a Mapping Ruleset, click **Add Rule** to define a new rule. Assume that you want to define a rule for all hosts that satisfy this set of conditions:

- IP address starts with `10.10.42`, AND:
- More than 6 CPUs OR `CRIBL_HOME` environment variable contains `DMZ`, AND:
- Belongs to `Fleet420`.

Rule Configuration

- **Rule Name**: `myFirstRule`

- **Filter**: `(conn_ip.startsWith('10.10.42.') && cpus > 6) || env.CRIBL_HOME.match('DMZ')`
- **Fleet**: `Fleet420`

Default Fleet and Mapping

When an Edge instance runs as Leader, the following are created automatically:

- A `default_fleet` Fleet.
- A `default` Mapping Ruleset,

Mapping Order of Priority

Priority for mapping to a fleet is as follows: Mapping Rules > Fleet (`cribl.group`) sent by Edge Node > `default_fleet` Fleet.

- If a Filter matches, use that Fleet.
- Else, if an Edge Node has a Fleet defined, use that.
- Else, map to the `default_fleet` Fleet.

Deploying Configurations

A typical workflow for deploying Edge configurations looks like this:

1. Work on configs.
2. Save your changes.
3. Commit (and optionally push).
4. Locate your desired Fleet and click **Deploy**.

Once you save and commit your configuration changes, you can deploy the updated configs to the Edge Nodes. You deploy new configurations at the Fleet level.

Deploying with Subfleets

On Cribl Edge v.4.1.x and later, when you commit changes to a parent Fleet, you have the option to **Deploy with Subfleets**.

When you **Deploy with Subfleets**, the Subfleets will receive the new inherited configuration.

Deploy with Subfleets

Edge Nodes that belong to the Fleet/Subfleet will start **pulling** updated configurations on their next check-in with the Leader.

Can't log in to the Edge Node as Admin User? When an Edge Node pulls its first configs, the admin password will be randomized, unless specifically changed. This means that users won't be able to log in on the Edge Node with default credentials. For details, see User Authentication.

Configuration Files

On the Leader, a Fleet's configuration resides under:

`$CRIBL_HOME/groups/<FleetName>/local/edge`.

On the managed Edge Node, after configs have been pulled, they are extracted under:

`$CRIBL_HOME/local/edge/`.

Some configuration changes will require restarts, while many others will require only reloads. For details, see Configurations and Restart. Upon restart, individual Edge Nodes might temporarily disappear from the Leader's **Nodes** tab, before reappearing.

Environment Variables

- `CRIBL_DIST_MASTER_URL` – URL of the Leader Node.
  Format: `<tls|tcp>://<authToken>@host:port?group=default_fleet&tag=tag1&tag=tag2&tls.<tls-settings below>`.
  Example: `CRIBL_DIST_MASTER_URL=tls://<authToken>@leader:4200`
  Parameters:
  - `group` – The preferred Fleet assignment.
  - `resiliency` – The preferred Leader failover mode.

    - `volume` – The location of the NFS directory to support Leader failover.
    - `tag` – A list of tags that you can use to assign the Edge Node to a Fleet.
    - `tls.privKeyPath` – Private Key Path.
    - `tls.passphrase` – Key Passphrase.
    - `tls.caPath` – CA Certificate Path.
    - `tls.certPath` – Certificate Path.
    - `tls.rejectUnauthorized` – Validate Client Certs. Boolean, defaults to `false`.
    - `tls.requestCert` – Authenticate Client (mutual auth). Boolean, defaults to `false`.
    - `tls.commonNameRegex` – Regex matching peer certificate > subject > common names allowed to connect. Used only if `tls.requestCert` is set to `true`.
- `CRIBL_DIST_MODE` – Set to one of: `managed-edge` (managed Node), `master` (Leader instance), or `edge` (single instance).
- `CRIBL_HOME` – Auto setup on startup. Defaults to parent of `bin` directory.
- `CRIBL_CONF_DIR` – Auto setup on startup. Defaults to parent of `bin` directory.
- `CRIBL_NOAUTH` – Disables authentication. Careful here!!
- `CRIBL_DIST_LEADER_BUNDLE_URL` – AWS S3 bucket (format: `s3://${bucket}`) for remote bundle storage.
- `CRIBL_TMP_DIR` – Defines the root of a temporary directory.
  Sources use this directory to stage downloaded Parquet data. The format of derived temporary directories is: `$CRIBL_TMP_DIR || os.tmpdir())/cribl_temp[/<componentPath>/<PID>]` ... where: `<componentPath>` = `<inputId>`; `<PID>` = Worker Process ID; and `os.tmpdir()` defaults to `/tmp` on Linux. Override that default using this `CRIBL_TMP_DIR` environment variable.
- `CRIBL_VOLUME_DIR` – Sets a directory that persists modified data between different containers or ephemeral instances.
- `CRIBL_DIST_WORKER_PROXY` - Communicate to the Leader Node via a SOCKS proxy. Format: `<socks4|socks5>://<username>:<password>@<host>:<port>`. Only `<host>:<port>` are required.
  The default protocol is `socks5://`, but you can specify `socks4://proxyhost:port` if needed.
  To authenticate on a SOCKS4 proxy with username and password, use this format: `username:password@proxyhost:port`. The `proxyhost` can be a `hostname`, `ip4`, or `ip6`.

Edge Node GUID

When you install and first run the software, a GUID is generated and stored in a `.dat` file located in `CRIBL_HOME/bin/`, e.g.:

```
# cat CRIBL_HOME/bin/676f6174733432.dat

{"it":1570724418,"phf":0,"guid":"48f7b21a-0c03-45e0-a699-01e0b7a1e061"}
```

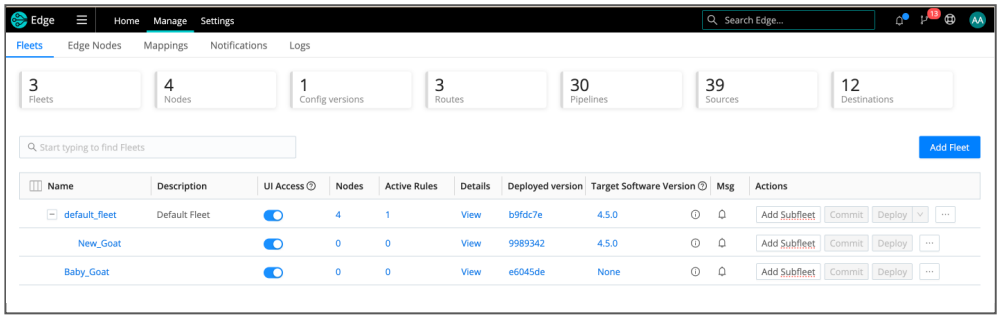
When deploying Cribl Edge as part of a host image or VM, be sure to remove this file, so that you don't end up with duplicate GUIDs. The file will be regenerated on next run.

4.2. Managing Edge Nodes

If you have an Enterprise or Standard license, you can click the **Manage** tab in the left nav to open the **Manage Fleets** page. This page has three upper tabs: **Fleets**, **Edge Nodes** and **Mappings**.

Fleets Tab

The **Manage Fleets** page provides a list of all configured Fleets and Subfleets in the instance.

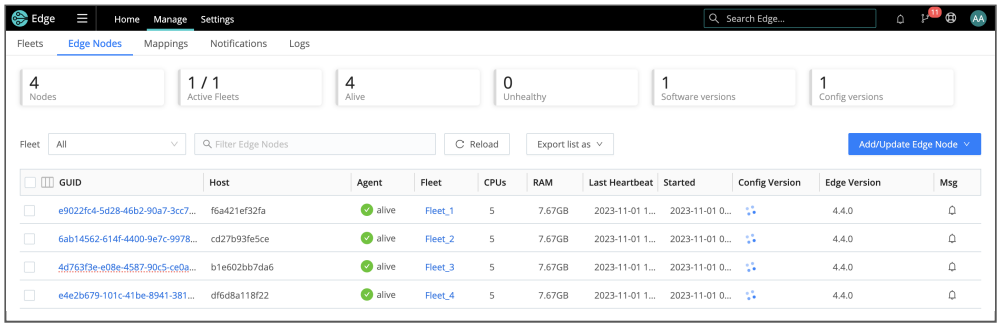


Manage Fleets

The top header now shows you the number of configured Fleets (2), along with other statistics. Clicking on the Fleet's **Name** redirects you to its Fleet Landing Page where you can explore more information for each of your configured Edge Nodes.

Edge Nodes Tab

The **Edge Nodes** tab on the **Manage Edge Nodes** page provides status information for each Edge Node in the selected Fleet.



Edge Node tab

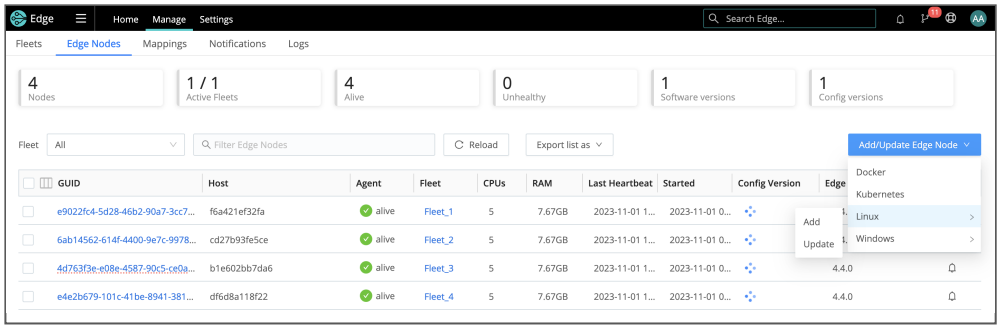
- To export the list of Edge Nodes, click the **Export list as** drop-down. Select `JSON` or `CSV` as the export file format. If you filter the list before exporting, the exported list will also be filtered. You can export the entire list by removing all filters.

- To display additional details and controls, click each row.
- To teleport from the Leader into the Edge Node, click the **Edge Node GUID** link.

Add/Update Edge Nodes

You can use the **Add/Update Edge Node** control at upper right to update an existing Node, or to add a new Node by generating a bootstrap script. Cribl Edge admins can use the UI to concatenate and copy/paste the bootstrap script, automating several steps below. You can use an adjacent option to grab a script that updates an Edge Node's Fleet assignment. You also have the option of generating a bootstrap script for the following Edge Nodes:

- Docker: For details, see Running in a Docker Container.
- Kubernetes: For details, see Deploying via Kubernetes.
- Linux: For details, see Installing Cribl Edge on Linux.
- Windows: For details, see Installing Cribl Edge on Windows.



Options to add or update an Edge Node

The **Update** option in the UI adjusts the Leader connection details for the currently installed software. To upgrade the Edge Node – that is, install a newer version of the software – see Upgrading.

Add/Bootstrap a New Edge Node

All Edge Nodes' hosts must enable ongoing outbound communication to the Leader's port 4200, to enable the Leader to manage the Nodes. While the bootstrap script runs, firewalls on each Nodes's host must also allow outbound communication on the following ports:

- Port 443 to `https://cdn.cribl.io`.
- Port 443 to a Cribl.Cloud Leader.
- Port 9000 to an on-premises Leader.

The details below differ slightly depending on which deployment option you select.

1. From Cribl Edges's top nav, select Manage > Edge Nodes.
2. On the resulting Edge Nodes tab, click **Add/Update Edge Node** at the upper right.
3. Select the deployment option as shown in the composite screenshots below.
4. In the resulting modal, the **Install package location** defaults to `Cribl CDN`. If desired, change this to `Download URL`.
5. As needed, correct the target **Fleet**, as well as the **Leader hostname/IP** (URI).
6. As needed, correct the **User** and **User Group** to run Cribl as. Defaults to `cribl`.
7. As needed, correct the **Installation directory**. Defaults to `/opt/cribl`.
8. Optionally, add **Tags** that you can use for filtering and grouping in Cribl Edge. Use a tab or hard return between (arbitrary) tag names.
9. Copy the resulting script to your clipboard.
10. Click either **Done** or **X** to close the modal.

Paste the script onto your Edge Node's command line and execute it, to add the Edge Node.

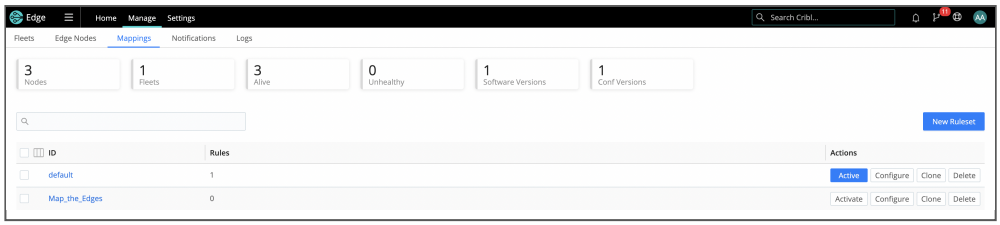
Troubleshooting the Display

If you see unexpected results on the **Edge Nodes** tab, keep in mind that:

- Edge Nodes that miss 5 heartbeats, or whose connections have closed for more than 30 seconds, will be removed from the list.
- For a newly created Fleet, the **Config Version** column can show an indefinitely spinning progress spinner for that Fleet. This happens because the Edge Nodes are polling for a config bundle that has not yet been deployed. To resolve this, click the **Deploy** option to force a deploy.
- For details on collocating multiple Cribl products, see Cribl Edge and Cribl Stream on the Same Host.
- For details on overriding default ports, see Overriding Default Ports.

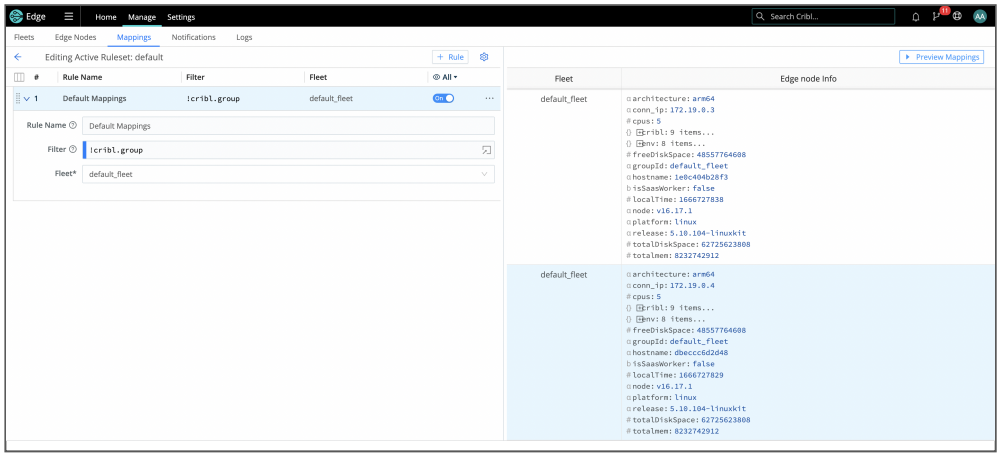
Mappings Tab

Click the **Mappings** tab (**Manage** > **Mappings**) to display status and controls for the active Mapping Ruleset. This page displays a maximum of 10,000 results.



Mappings status/controls

To manage and preview the Rules in a Ruleset, click into it.



Managing Ruleset page

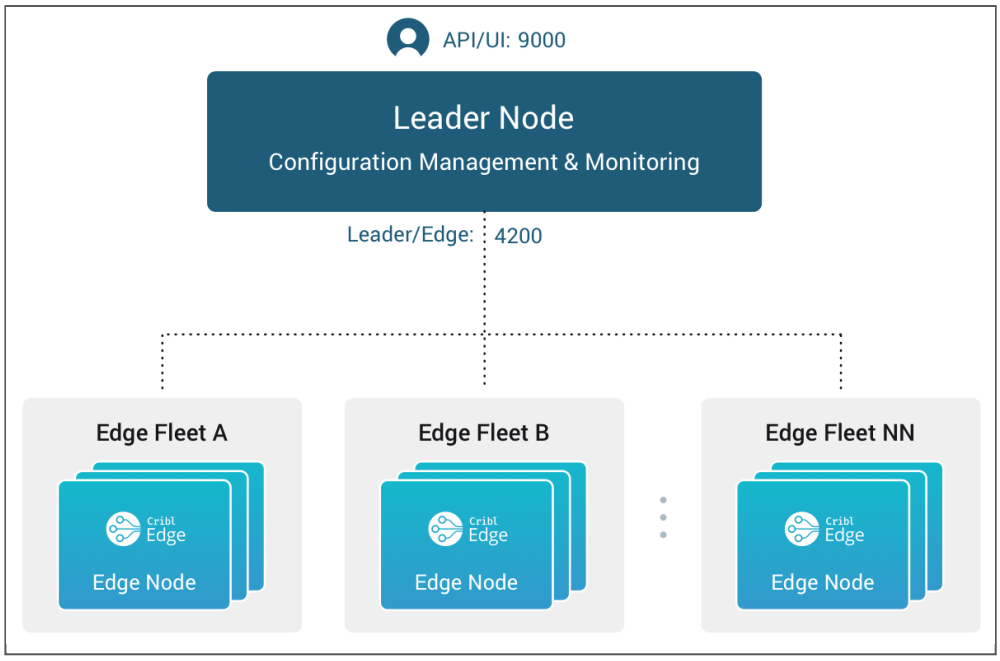
Collecting AWS EC2 Tag Metadata

If Edge is connected to an AWS EC2 instance, the **Mappings** tab can gather metadata for AWS instance tags. You can use tags to map EC2 nodes to the correct Edge Fleet. You must allow access to tags in the EC2 instance metadata for Cribl Edge to obtain them.

How Edge Nodes and Leader Work Together

The Leader Node has two primary roles:

1. Serves as a central location for Edge Nodes' operational metrics. The Leader ships with a monitoring console that has a number of dashboards, covering almost every operational aspect of the deployment.
2. Serves as a central location for authoring, validating, deploying, and synchronizing configurations across Fleets.



Leader Node/Edge Nodes relationship

Network Port Requirements (Defaults)

- UI access to Leader Node: TCP 9000.
- Edge Node to Leader Node: TCP 4200. Used for Heartbeat/Metrics/Leader requests/notifications to clients (for example: live captures, teleporting, status updates, config bundle notifications, and so on).
- Edge Node to Leader Node: HTTPS 4200. Used for config bundle downloads.
- Edge Node boostrapping/update: TCP 443.

Leader/Edge Node Communication

Edge Nodes will send a heartbeat to the Leader every 60 seconds. This heartbeat includes information about the Nodes and a set of current system metrics. The heartbeat payload includes facts – such as hostname, IP address, GUID, tags, environment variables, current software/configuration version, etc. – that the Leader tracks with the connection.

The failure of an Edge Node to successfully send two consecutive heartbeat messages to the Leader will cause the respective Edge Nodes to be removed from the Nodes page in the Leader's UI until the Leader receives a heartbeat message from the affected node.

When an Edge Node checks in with the Leader:

- It sends a heartbeat to Leader with "facts" about itself.
- The Leader uses Edge Node's facts and Mapping Rules to map it to a Fleet.
- The Edge Node pulls its Fleet's updated configuration bundle, if necessary.

The Leader is unaware of Edge Nodes' platforms (i.e, Linux or Windows) within a Fleet. So the ConfigHelper omits platform-specific limitations. Therefore, when you manage Edge Nodes on heterogeneous platforms, create a Windows-specific Fleet and mapping. See Managing Edge Nodes on Multiple Platforms.

Config Bundle Management

Config bundles are compressed archives of all config files and associated data that an Edge Node needs to operate. The Leader creates bundles upon Deploy, and manages them as follows:

- Bundles are wiped clean on startup.
- While running, at most 5 bundles per group are kept.
- Bundle cleanup is invoked when a new bundle is created.

The Edge Node pulls bundles from the Leader and manages them as follows:

- Last 5 bundles and backup files are kept.
- At any point in time, all files created in the last 10 minutes are kept.
- Bundle cleanup is invoked after a reconfigure.

Bundle in S3

You can store bundles in an AWS S3 bucket to reduce the Leader load. This offloads bundle distribution from the Leader to your designated S3 bucket. Edge Nodes will pull bundles directly from S3, minimizing Leader strain.

You can configure bundling in S3 through the following ways:

- **UI**: When configuring **Leader Settings** on a Leader Node, specify an S3 bucket (format: `s3://${bucket}`) for remote bundle storage in the **S3 Bundle Bucket URL** field.
- **YAML**: In the `instance.yml` config file, specify the S3 bucket in `master.configBundles.remoteUrl`.
- **Environment Variable**: Use the `CRIBL_DIST_LEADER_BUNDLE_URL` environment variable.

4.3. Setting Up Leader And Edge Nodes

This page covers:

1. Configuring a Leader Node
2. Configuring an Edge Node

Configuring a Leader Node

You can configure a Leader Node either through the UI or through the `instance.yml` config file.

Using the UI

In **Settings** (top nav) > **Global Settings** > **Distributed Settings** > **Distributed Management** > **General Settings**, select **Mode**: **Leader**.

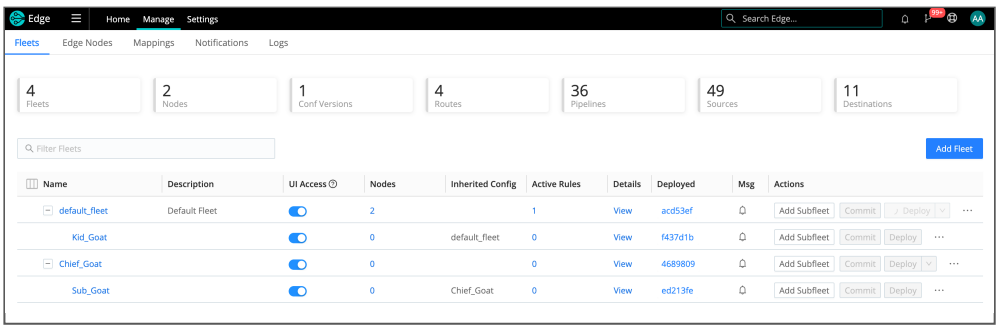
Next, on the **Leader Settings** left tab, confirm or enter the required Leader settings (**Address** and **Port**). Customize the optional settings if desired. Then click **Save** to restart.

Edge Node UI Access

This useful option enables you to click through from the Leader's **Manage Edge Nodes** page to an authenticated view of each Node's UI. The instructions below correspond to enabling the `groups.yml` file's `workerRemoteAccess` configuration key.

To enable Node UI access from the Leader's UI:

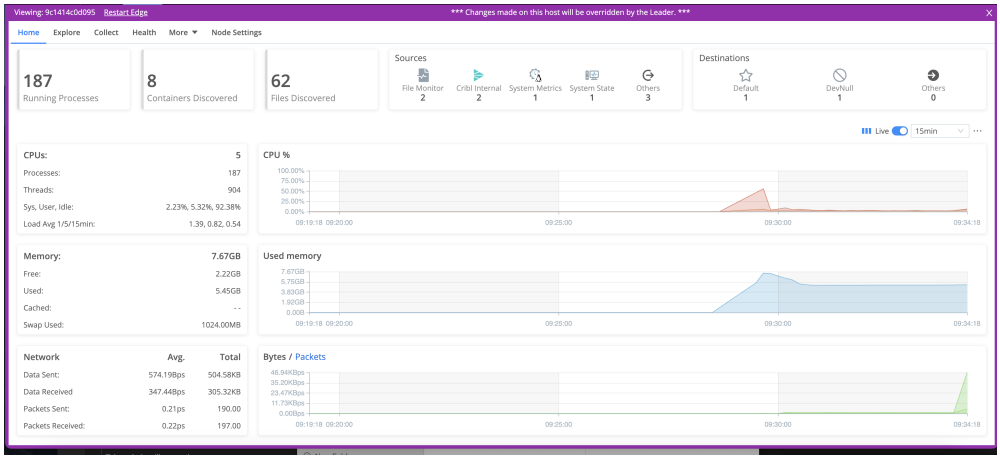
1. From Cribl Edge's top nav, select **Manage**.
2. On the **Manage Fleets** page: For each desired Fleet, toggle **UI Access** to `On`.



Manage Fleets

3. From the top nav, select **Edge Nodes**.

4. On the **Manage Edge Nodes** page, click the link for any Node you want to inspect.
   To confirm that you are remotely viewing a Edge Node's UI, Cribl Edge displays a purple border, with a badge labeled **Viewing host:** `<host/GUID>`.



Authenticated view of an Edge Node

The Leader will override any changes that you make directly to this Edge Node.

Using YAML Config File

In `$CRIBL_HOME/local/_system/instance.yml` (`C:\Program Data\Cribl\local\_system.yml` on Windows), under the `distributed` section, set `mode` to `master`:

$CRIBL_HOME/local/_system/instance.yml or C:\Program Data\Cribl\local_system.yml

```
distributed:
  mode: master
  master:
    host: <IP or 0.0.0.0>
    port: 4200
    tls:
      disabled: true
    ipWhitelistRegex: /.&ast;/
    authToken: <auth token>
    enabledWorkerRemoteAccess: false
    compression: none
    connectionTimeout: 5000
    writeTimeout: 10000
```

Persisting Socket Connections

A distributed deployment creates socket files for inter-process communication (IPC) between the Leader and distributed processes and services. These sockets are essential for ensuring that Edge Nodes successfully connect to the Leader, and for certain metrics services. On the Leader's host, the default location for these files is the operating system's temp directory (e.g., `/tmp`).

Many Linux distros maintain a system cleaner service (e.g., systemd-tmpfiles) that removes files from this directory periodically, such as every 10 days. If Cribl's sockets are removed, this breaks certain UI pages, such as those for Fleets and Monitoring. You can protect the sockets in either of two ways.

Block the Cleaner

Stop the host OS from cleaning socket files out of `/tmp/cribl-*` subdirectories. For example, on Amazon Linux 2 instances, add a new file to `/etc/tmpfiles.d` with the line: `X /tmp/cribl-*`.

To restart the system cleaner here and reload its configuration, use this command:
`systemctl restart systemd-tmpfiles-clean.service`

Move the Sockets

Alternatively, you can move Cribl's socket files to a different directory. This directory must be outside your operating system's temp directory, and relatively close to the root; and the Cribl admin must have user permissions to write to it. As one example of a protected directory, you could specify: `/var/tmp`

In the UI, you specify the directory at **Settings** > **Global Settings** > **System** > **Distributed Settings** > **Leader Settings** > **Helper processes socket dir**.

Configuring an Edge Node

On each endpoint, you can configure Cribl Edge variously through the UI, the `instance.yml` config file, environment variables, or the command line.

Using the UI

In **Settings** (top nav) > **Global Settings** > **Distributed Settings** > **Distributed Management** > **General Settings**, select **Mode**: `Managed Edge`.

Next, on the **Leader Settings** left tab, confirm or enter the required **Address** (e.g., `criblleader.mycompany.com`).

Customize the optional settings if desired. Then click **Save** to restart.

Using YAML Config File

In `$CRIBL_HOME/local/_system/instance.yml` (`C:\Program Data\Cribl\local\_system.yml` on Windows), under the `distributed` section, set `mode` to `managed-edge`:

$CRIBL_HOME/local/_system/instance.yml or C:\Program Data\Cribl\local_system.yml

```
distributed:
  mode: managed-edge
  envRegex: /^CRIBL_/
  master:
    host: <master address>
    port: 4200
    authToken: <token here>
    compression: none
    tls:
      disabled: true
    connectionTimeout: 5000
    writeTimeout: 10000
  tags:
      - tag1
      - tag2
      - tag42
  group: teamsters
```

Using Environment Variables

You can configure Edge Nodes via environment variables, as in this example:

```
CRIBL_DIST_MASTER_URL=tcp://${CRIBL_DIST_TOKEN:-criblmaster}@masterHostname:4203
./cribl start
```

For additional details, see Environment Variables.

Using the Command Line

You can configure an Edge Node using CLI commands of this form:

```
./cribl mode-managed-edge -H <master-hostname-or-IP> -p <port> [options] [args]
```

The `-H` and `-p` parameters are required. For other options, see the CLI Reference. Here is an example command:

```
./cribl mode-managed-edge -H 192.0.2.1 -p 4200 -u myAuthToken
```

Edge will need to restart after this command is issued.

4.4. Leader High Availability/Failover

To handle unexpected outages in your on-premises distributed deployment, Cribl Edge 3.5 and above supports configuring a second Leader for failover. This way, if the primary Leader goes down, Collectors and Collector-based Sources can continue ingesting data without interruption.

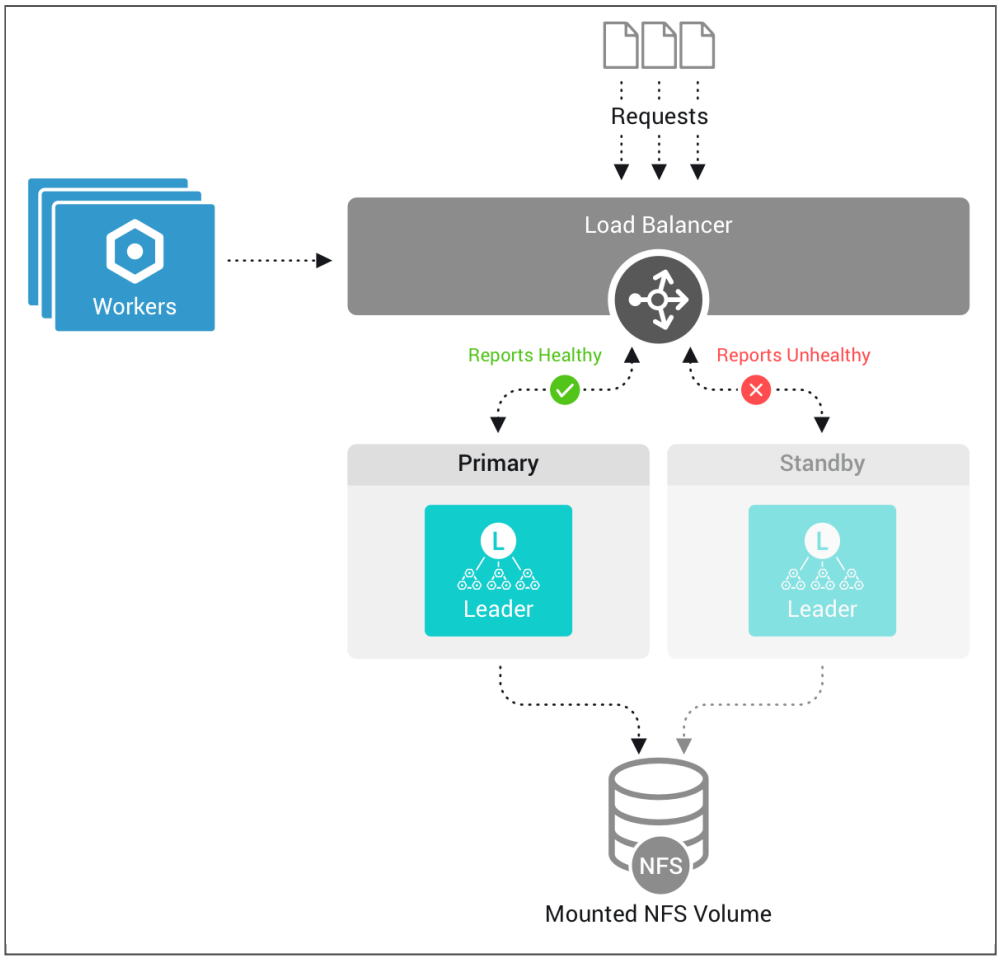
Configuring a backup Leader requires a Cribl Stream Enterprise or Standard license.

How It Works

When you configure a second Leader, there will be only one active Leader Node at a time. The second Leader Node will be used only for failover. For this architecture to work, you must configure all failover Leaders' volumes to point at the same Network File System (NFS) volume/shared drive.

If the primary Leader Node goes down:

- Cribl Edge will recover by switching to the standby Leader Node.
- The new Leader Node will have the same configs, state, and metrics as the previous Leader Node.
- The Edge Nodes connect to the new Leader.



Leader High Availability/Failover Design

Required Configuration

Before adding a second Leader, ensure that you have the configuration outlined in this section.

Auth Tokens

Make sure that both Leaders have matching auth tokens. If you configure a custom **Auth token**, match its value on the opposite Leader.

- In the UI, check and match these values at each Leader's **Settings** > **Global Settings** > **Distributed Settings** > **Leader Settings** > **Auth token**.
- Or, from the filesystem, check and match all Leaders' instance.yml > `master` section > `authToken` values.

(If tokens don't match, Edge Nodes will fail to authenticate to the alternate Leader when it becomes active.)

NFS

- On all Leader Nodes, use the latest version of the NFS client. **NFSv4 is required.**
- Ensure that the NFS volume has at least 100 GB available disk space.
- Ensure that the NFS volume's IOPS (Input/Output Operations per Second) is ≥ 200. (Lower IOPS values can cause excessive latency.)
- Ensure that ping/latency between the Leader Nodes and NFS is < 50 ms.

You can validate the NFS latency using a tool like `ioping`. Navigate to the NFS mount, and enter the following command:

```
ioping .
```

For details on this particular option, see the ioping docs.

NFS Mount Options

The Leader Node will access large numbers of files whenever you use the UI or deploy configurations to Cribl Edge Edge Nodes. When this happens, NFS's default behavior is to synchronize access time updates for those files, often across multiple availability zones and/or regions. To avoid the problematic latency that this can introduce, Cribl recommends that you add one of the following NFS mount options:

1. `relatime`: Update the access time only if it is more than 24 hours ago, or if the file is being created or modified. This allows you to track general file usage without introducing significant latency in Cribl Edge. To do the same for folders, add the `reldiratime` option.
2. `noatime`: Never update the access time. (Cribl Edge does not need access times to be updated to operate correctly.) This is the most performant option – but you will be unable to see which files are being accessed. To do the same for folders, add the `nodiratime` option.

Load Balancers

- Configure all Leaders behind a load balancer.
- Expose ports `9000` and `4200` via the load balancer.
- Load balancers must support health checks via `/api/v1/health` endpoint.

The following load balancers support health checks:

- Amazon Web Services (AWS) Network Load Balancer (NLB). Suitable for TCP, UDP, and TLS traffic.
- AWS Application Load Balancer (ALB). Application-aware, suitable for HTTP/HTTPS traffic.
- HAProxy.
- NGINX Plus.

AWS Network Load Balancers

If you need to access the same target through a Network Load Balancer, use an IP-based target group and deactivate client IP preservation. For details, see:

- Why can an instance in a target group not reach itself via NLB?
- Why can't a target behind my Network Load Balancer connect to its own Network Load Balancer?

Recommended Configuration

Use the latest NFS client across all Leaders. If you are on AWS, we recommend the following:

- Use Amazon's Elastic File System (AWS EFS) for your NFS storage.
- Ensure that the user running Cribl Edge has read/write access to the mount point.
- Configure the EFS **Throughput mode** to `Enhanced` > `Elastic`.
- For details on NFS mount options, see Recommended NFS mount options.

For best performance, place your Leader Nodes in the same geographic region as the NFS storage. If the Leader and NFS are distant from each other, you might run into the following issues:

- Latency in UI and/or API access.
- Missing metrics between Leader restarts.
- Slower performance on data Collectors.

Set the primary Leader's **Resiliency** drop-down to `Failover`.

Configuring Additional Leader Nodes

You can configure additional Leader Nodes in the following ways. These configuration options are similar to configuring the primary Leader Node:

- Using the UI
- Updating the YAML config file
- Using the Command Line
- Using Environment Variables

Remember, the `$CRIBL_VOLUME_DIR` environment variable overrides `$CRIBL_HOME`.

Using the UI

1. In **Settings** > **Global Settings** > **Distributed Settings** > **General Settings**, select **Mode**: `Leader`.
2. Next, on the **Leader Settings** left tab, select **Resiliency**: `Failover`. This exposes several additional fields.
3. In the **Failover volume** field, enter the NFS directory to support Leader failover (e.g., `/mnt/cribl` or `/opt/cribl-ha`). Specify an NFS directory outside of `$CRIBL_HOME`. A valid solution is to use `CRIBL_DIST_MASTER_FAILOVER_VOLUME=<shared_dir>`. See Using Environment Variables for more information.
4. Optionally, adjust the **Lease refresh period** from its default `5s`. This setting determines how often the primary Leader refreshes its hold on the Lease file.
5. Optionally, adjust the **Missed refresh limit** from its default `3`. This setting determines how many Lease refresh periods elapse before standby Nodes attempt to promote themselves to primary.
6. Click **Save** to restart.

In Cribl Edge 4.0.3 and later, when you save the **Resiliency**: `Failover` setting, further **Distributed Settings** changes via the UI will be locked on both the primary and backup Leader. (This prevents errors in bootstrapping Workers due to incomplete token synchronization between the two leaders.) However, you can still update each Leader's distributed settings by modifying its configuration files, as covered in the very next section.

Using the YAML Config File

In `$CRIBL_HOME/local/_system/instance.yml`, under the `distributed` section:

1. Set `resiliency` to `failover`.
2. Specify a volume for the NFS disk to automatically add to the Leader Failover cluster.

$CRIBL_HOME/local/_system/instance.yml

```
distributed:
  mode: master
  master:
    host: <IP or 0.0.0.0>
    port: 4200
    resiliency: failover
    failover:
      volume: /path/to/nfs
```

Note that `instance.yml` configs are local, not on the shared NFS volume.

Using the Command Line

You can configure another Leader Node using a CLI command of this form:

```
./cribl mode-master -r failover -v /tmp/shared
```

For all options, see the CLI Reference.

Using Environment Variables

You can configure additional Leader Nodes via the following environment variables:

- `CRIBL_DIST_MASTER_RESILIENCY=failover`: Sets the Leader's `Resiliency` to `Failover` mode.
- `CRIBL_DIST_MASTER_FAILOVER_VOLUME=<shared_dir>`: Sets the location (e.g., `/mnt/cribl`) of the NFS directory to support Leader failover.
- `CRIBL_DIST_MASTER_FAILOVER_MISSED_HB_LIMIT`: Determines how many Lease refresh periods elapse before the standby Nodes attempt to promote themselves to primary. Cribl recommends setting this to `3`.
- `CRIBL_DIST_MASTER_FAILOVER_PERIOD`: Determines how often the primary Leader refreshes its hold on the Lease file. Cribl recommends setting this to `5s`.

For further variables, see Environment Variables.

Monitoring the Leader Nodes

To view the status of your Leader Nodes, select **Monitoring** > **System** > **Leaders**.

Monitoring Leader Nodes

Upgrading

When upgrading:

1. Stop both Leaders.
2. Upgrade (Stream, Edge) the primary Leader.
3. Upgrade the second Leader.
4. Start both Leaders again, one by one.
5. Upgrade each Edge Node, respectively.

Disabling the Second Leader

Cribl recommends that you maintain a second Leader to ensure continuity in your on-premises distributed environment. Should you decide to disable it, contact support to assist you.

5. Cribl Edge Better Practices

5.1. Monitoring Your Infrastructure With Cribl Edge

Cribl Edge helps you monitor your sprawling infrastructure by collecting, processing, and forwarding data with low overhead cost and resources. Cribl Edge simplifies agent administration and data collection for your endpoints, and provides a number of charts, graphs, and alerts for the web when connected to supported data sources.

To enhance the performance of your environments, combine Cribl Edge with your data visualization tool of choice (this guide uses Grafana as an example). Now, you can identify and investigate issues and monitor applications faster than ever before.

Useful Metrics Dashboards

Cribl has configured a Prometheus dashboard to display system metrics, which is available in Grafana's public library. This is a relatively simple dashboard, suitable for showing basic aggregate metrics.

Another useful dashboard for displaying Linux metrics is the Node Exporter Full dashboard, commonly used with Prometheus and its Node Exporter agent. This dashboard can handle highly detailed metrics.

To display Windows metrics, consider one of Grafana's Windows dashboards.

Guide Overview

This setup guide will walk you through these configuration steps:

1. Configure your Cribl Edge Sources. Depending on your endpoints, you can use the System Metrics Source for Linux, Windows Metrics, or Kubernetes Metrics.
2. Configure the Grafana Cloud Destination to send data to Grafana Cloud.
3. Access Grafana Cloud to see the metrics and visualizations.

Configure the System Metrics Source on Cribl Edge

1. In Cribl Edge, start by configuring and enabling the applicable Source for system metrics:

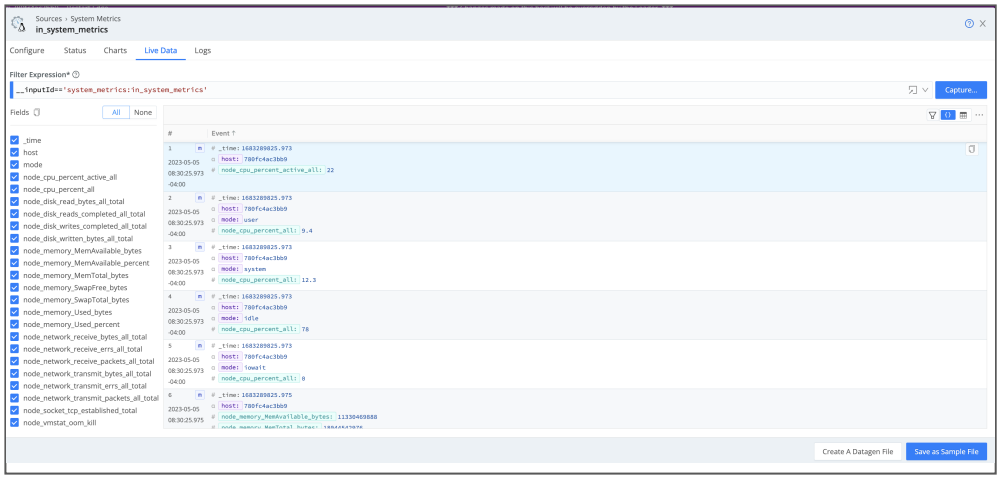
    - Linux System Metrics Source
    - Windows Metrics
    - Kubernetes Metrics
2. Decide what level of metrics detail you want to collect and display.

In the System Metrics Source and Windows Metrics Sources, go to the **Host Metrics** and **Container Metrics** tabs:

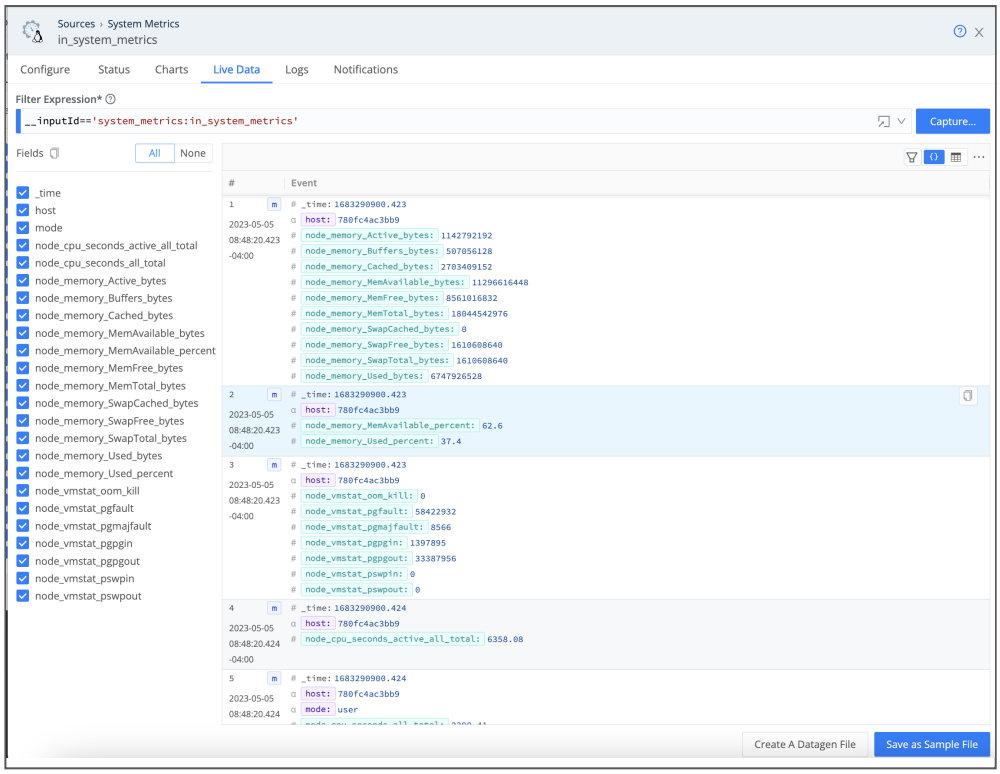
- For the relatively simple dashboard showing aggregate metrics, choose the **Basic** mode for most of your metrics.
- For the highly detailed metrics visualizations dashboard, choose the **All** mode for all of your metrics.

For the Kubernetes Metrics Source, use the **Filter Rules** to specify the Kubernetes objects that the Source should parse to generate metrics. If you specify no restrictive rules here, the Source will emit all events.

1. For all of the Sources, click **Processing Settings > Pre-Processing**, then select the `prometheus_metrics` **Pipeline**.
2. When you're finished, **Commit** and **Deploy** the changes. Before you move on to the next step, check the **Live Data** tab to make sure the metrics are generated.



Basic System Metrics

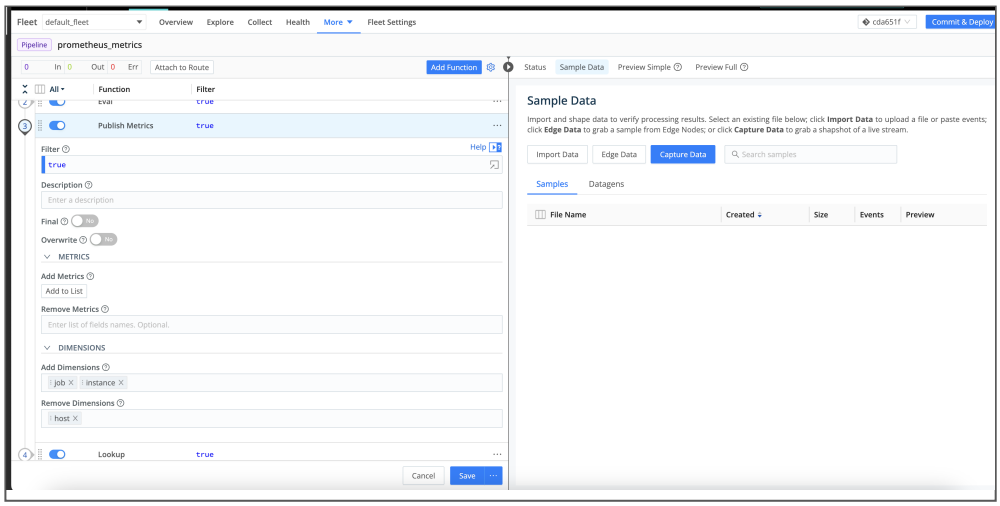


Detailed System Metrics

(Optional) Adjust the `prometheus_metrics` Pipeline

Optionally, you can adjust the `prometheus_metrics` Pipeline to include the `host` dimension, or to add/remove more dimensions.

Select **Manage** from the top nav, then select a **Fleet** to configure. From a Fleet's top nav, select **More** > **Pipelines**.

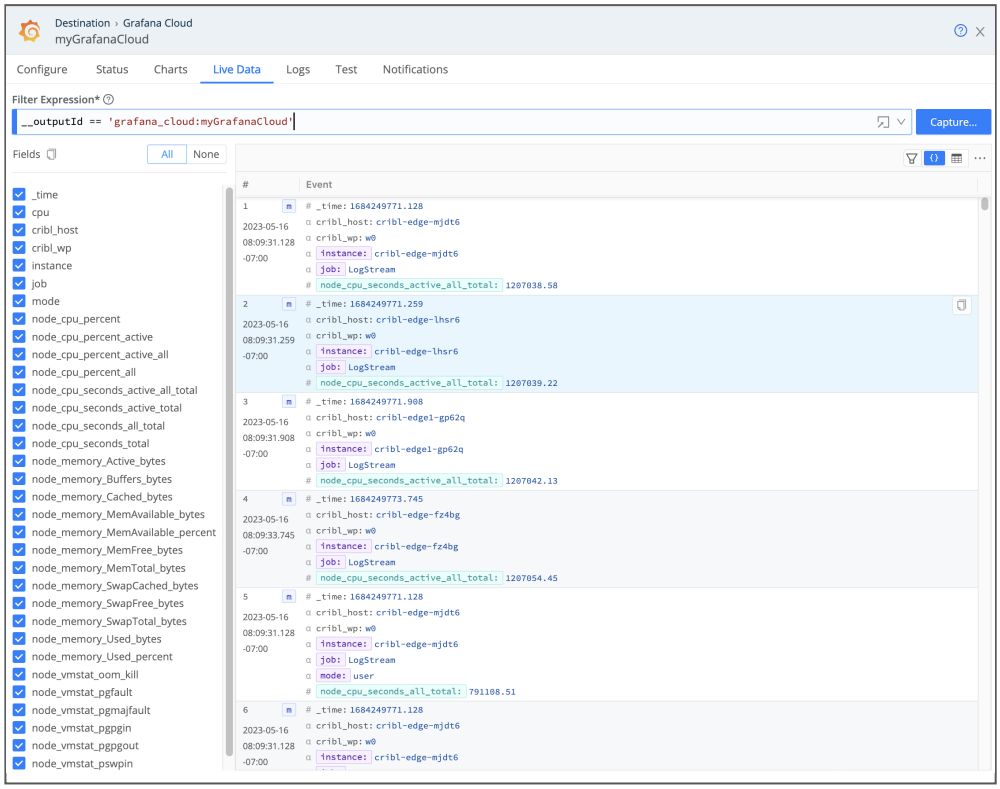


Customizing the Pipeline's dimensions

Configure the Grafana Cloud Destination

Next, head over to Grafana Cloud Destination to configure the Destination and send data to Grafana Cloud.

Preview your data on the Grafana Cloud config modal's **Live Data** tab.

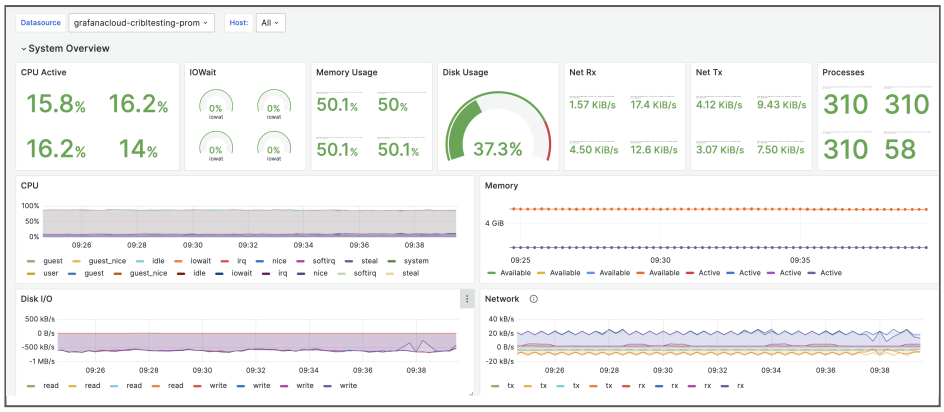


Preview Live Data

Visualize the Data in Grafana

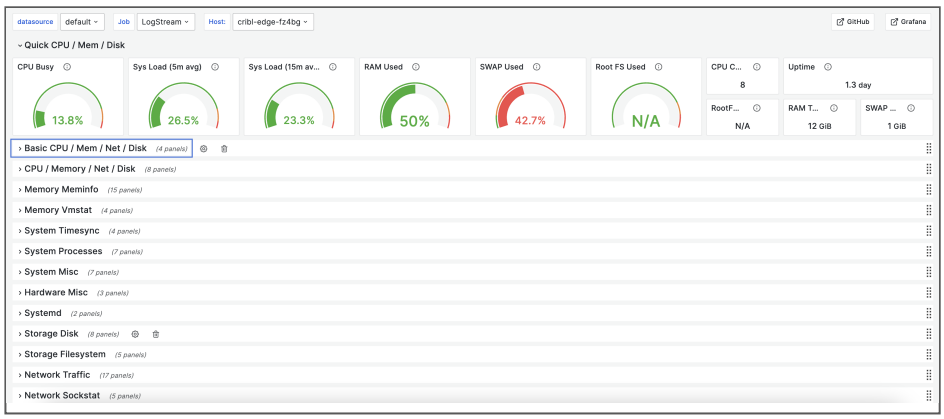
You can now visualize metrics in Grafana Cloud. To get started:

1. Navigate to your Grafana Cloud instance, and select **Explore** on the left panel to review the Node metrics.
   You can select metrics and run a query to produce a chart and table of metric events, which you can add to new or existing dashboards for analysis. If expected metrics are missing, try adjusting the variables in your Grafana Cloud instance. See Grafana's Variables topic for details.
2. Use Grafana's **Filter** option to browse the dashboards for your Prometheus data. Below are sample dashboards with the basic and detailed metrics shown.
   - To show basic aggregate metrics, use the Cribl Node Observability dashboard.



Cribl Node observability

- For the `OS` metrics with `CPU` / `Disk` details, the basic dashboard will not suffice. Instead, we recommend configuring your System Metrics Source (as discussed above) to receive **All** metrics, and then using the Node Exporter Full dashboard below for visualization.



Node Exporter Full

5.2. Fleets Hierarchy And Design

Cribl Edge enables you to flexibly group Edge Nodes into logical Fleets and Subfleets, organized as you like. However, Fleet design is extremely important for the health of the Leader. In this guide, we address "better practices" for designing your Fleets and Subfleets.

Leader Health

Below are some guidelines for designing your Fleets to optimize your Leader's health:

- Do not exceed 3,000 Edge Nodes per Fleet hierarchy, which is a Fleet and all of its Subfleets. Exceeding this maximum causes issues when the Leader pushes a configuration bundle. For details, see Leader Scalability.
- Consider the number of Edge Nodes restarting at once, since they will all be pulling their configurations from the Leader at the same time.
- For deployments with a larger number of Edge nodes, change the default metrics controls to `minimal`.
- And lastly, since upgrades occur at the Fleet level, consider the number of Edge Nodes the Leader will be upgrading at once.

Fleet Design Better Practices for Enterprises

Consider the following "better practices" for designing Fleets for enterprises:

- Do not use the `default_fleet`. The `default_fleet` is the fallback for mapping errors. If you purchased an Enterprise License, disable all the Sources in the `default_fleet` to prevent accidental collections. This guideline doesn't apply to the free license, where you are limited to the `default_fleet`.
- Design your Fleet hierarchy before deploying your Edge Nodes. For details, see Designing a Fleet Hierarchy below.
- Use the bootstrap script in the Add/Update Edge Node modal to deploy Edge Nodes to a specific Fleet. Depending on your Fleet design, we recommend adding tags to the bootstrap script to uniquely identify the Fleet. You can use these tags for automatic mappings.
- Create mappings for all the Fleets so that Edge Nodes are automatically placed in the Fleet when they bootstrap and connect to the Leader. This way, if any rogue Edge Node gets deployed, it will automatically show up in `default_fleet`. If you disable all the Sources in the `default_fleet`, there will be no impact on data consumption.
- Monitor your `default_fleet` for any unintentional/rogue Edge Nodes.

Designing Your Fleet Hierarchy

The first question you should ask is, do you need Subfleets? In some cases, a single-level (flat) design is sufficient.

In more complex environments, layering the configurations can help you to manage a large number of Edge Nodes that share common yet differentiated configurations. For example, as an administrator, you might need multiple layers of configurations to manage the following: `Org` > `Department` > `OS` > `Application`. In this case, the Edge Nodes at the `Application` level can inherit and share configurations from the `Org` level. You can update a parent-level configuration (Fleet) and have it applied to all Subfleets and their respective Edge Nodes, reducing the time and effort needed to manage them.

There are several ways you can build the hierarchy:

- Geographical: If your organization collects different data categories in different regions, consider this option.
- Data center: This is a good option if you maintain data centers in different regions.
- Server function (web server, database, etc.): Another good way to organize your Fleets can be based on server functions.
- OS-based: If you are running an environment with both Linux and Windows machines, consider separating them into different Fleets per OS.

The recommendations here:

- To avoid confusion, try to keep Fleets to two to three levels. We'll discuss it more under Inheritance Considerations.

To learn more about Fleets, see Fleets and check out Configuring a Fleet for configuration details.

Naming your Fleets

What's in a name, you ask? Well, a whole lot when it comes to Fleets. Consider naming your Fleets and Subfleets in a way that helps you recognize the configurations at a glance.

For example, you can use the top-level Fleet name as part of your lower-level Fleets. Naming your Fleet with a logical `<parent-fleet>-<subfleet1>-<subfleet2>` will easily provide you with context whether you are viewing at the Fleet, Subfleet, or Edge Node level. It can also help identify Edge Nodes that are out of place or mapped incorrectly.

Here's an example of naming a four-level Fleet in an intuitive way:

- BZT

    - BZT-NewYork
        - BZT-NY-Linux
            - BZT-NY-Linux-Apache
            - BZT-NY-Linux-Gizmotrax
        - BZT-NY-Windows
            - BZT-NY-Win-IIS
    - BZT-LA
        - BZT-LA-K8s
        - etc.

Other guidelines to keep in mind when naming your Fleets:

- Fleet names can't have spaces in them; use a dash or underscore instead.
- You can't reuse a Fleet name.
- Once you create a Fleet, you can't edit the name.
- You can't delete a Fleet if there are Subfleets under it.

Inheritance Considerations

As suggested above in Designing Your Fleet Hierarchy, plan and design the Fleets' inheritance in advance. Here are some guidelines to keep in mind, as you think about layering your configurations.

Top-level Fleets (Grandparents)

Assign the top-level Fleets for company-wide configurations. Consider:

- Applying certificates and other company-wide settings at this level.
- Avoid OS-specific or site-specific configurations at this level.

Subfleet 1 (Parents)

Per-site configurations. Consider:

- Configure site-specific Sources/Destinations at this level.
- Specify any site-specific configurations like common Destinations, etc.

Subfleet 2 (Children)

Configurations used by groups of similar systems. Consider:

- Apply OS-specific configurations at this level. For example, an “All Linux” and “All Windows” for a given site.
- Group common OS-specific metrics and common log files into OS-specific Fleets.

Subfleet 3 (Grandchildren)

Apply function-specific Subfleets to add specific metrics, logs, and/or Windows Event sources. For example, you can group by:

- Apache HTTPD
- Database
- SMTP servers
- Active Directory
- IIS
- Exchange

If you make a change to an inherited config (such as Sources/Destinations) in a Subfleet, the Subfleet will no longer get updates made to the parent Fleet for that inherited config. Other configs will still be inherited from that parent.

When you make changes to a parent Fleet configuration, use the **Commit and Deploy with Subfleets** option. This way, you are always keeping the inherited configurations in sync.

Finally, note that Packs and Pipelines can’t be inherited by a Subfleet from a Fleet at this time. They need to be recreated / imported into child fleets.

5.3. Kubernetes Observability Using Cribl Edge

Capturing observability data is difficult in Kubernetes environments. Kubernetes deployments can scale up to thousands of nodes, making monitoring challenging. In addition, tracking and troubleshooting infrastructure performance is challenging due to its dynamic nature.

In this guide, we will show you how to address these problems by:

- Deploying Cribl Edge via Kubernetes using the Cribl Helm Chart, including creating Kubernetes-specific Fleets and > Fleet settings.
- Setting environment variables for your Kubernetes environment to work.
- Configuring the Kubernetes Sources, Destinations, and Pipelines to connect them.
- Visualizing the data in Grafana using some pre-built dashboards.
- Learning about “better practices” and recommendations for deploying Cribl Edge on your Kubernetes clusters.

While some open-source tools can address these needs, they may be cumbersome to deploy and manage for sprawling microservices environments. Getting Kubernetes data into analytics platforms is challenging due to its volume, lack of standard formatting, and ephemeral nature.

In addition to addressing these issues, Cribl Edge brings Cribl Stream’s processing power for deployment on servers. With the ability to collect, process, and forward data with low resource overhead, Cribl Edge is designed for sprawling environments such as Kubernetes. With Cribl Edge, you can capture Pod logs and Control Plane events, generate Kube State Metrics (KSM), scrape data from Prometheus endpoints, and send the data to your downstream systems.

Deploy Cribl Edge via Kubernetes

In this section we’ll cover:

- Configuring Fleets and Fleet Settings
- Installing Using Helm
- Setting Environment Variables for your Kubernetes Deployment

Configuring Fleets & Fleet Settings on Cribl Edge

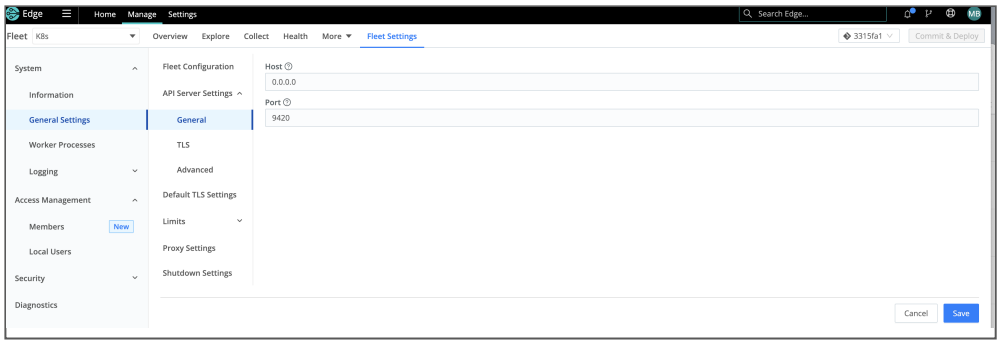
Before we deploy the Edge Node, let’s configure a Fleet. Fleets and their corresponding Subfleets allow you to author and manage configuration settings for a particular set of Edge Nodes.

We recommend configuring a separate Fleet to manage your Kubernetes environment. It is easier to deploy or upgrade Kubernetes Edge Nodes using our Helm Charts when they are all in one Fleet. For best practices on configuring Fleets, see Fleet Design and Hierarchy.

Once you have configured your Fleet, you must adjust the API Server settings in Fleet Settings to bind to all IPs in the container - this is done by changing API Server configuration from `127.0.0.1` (Loopback Interface) to `0.0.0.0`. Kubernetes health checks will fail if this step is not performed prior to pushing config bundles to Edge agents. For details, see Health Checks.

To change the Fleet Settings:

1. Navigate to **Manage** in the Cribl Edge UI.
2. Select the Kubernetes-specific Fleet.
3. Click **Fleet Settings** in the submenu.
4. Navigate to **System** > **General Settings** > **API Server Settings** > **General**.
5. Change the **Host** value from `127.0.0.1` to `0.0.0.0`.
6. Click **Save**.



Fleet Settings

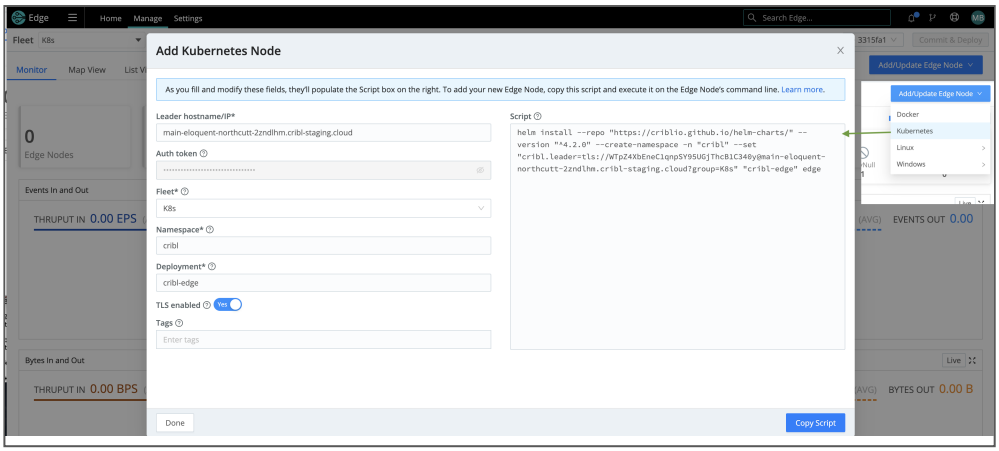
Use the Bootstrap Script to install the Helm Chart

We recommend deploying the Cribl Edge agent in your Kubernetes cluster using our Helm charts. These charts provide the minimum configuration required to install Cribl Edge correctly.

Our deployment has been tested across multiple Kubernetes platforms on several cloud service providers. The Cribl Helm Charts provide a fast, repeatable, and consistent way of deploying and upgrading an entire Kubernetes environment.

Next, bootstrap a new Edge Node on your Kubernetes Fleet using the Helm Script found in the UI. Copy/paste the bootstrap script for adding a Kubernetes node.

Below is a composite screenshot of adding an Edge Node via Kubernetes.



Generating a script to bootstrap an Edge Node on Windows

See Deploying via Kubernetes for details on installing Cribl Edge.

Environment Variables and Metadata

You can customize environment variables and metadata to modify (respectively) your Edge Nodes' deployment and reporting.

Environment Variables

You can use the following environment variables to configure Cribl Edge via Kubernetes:

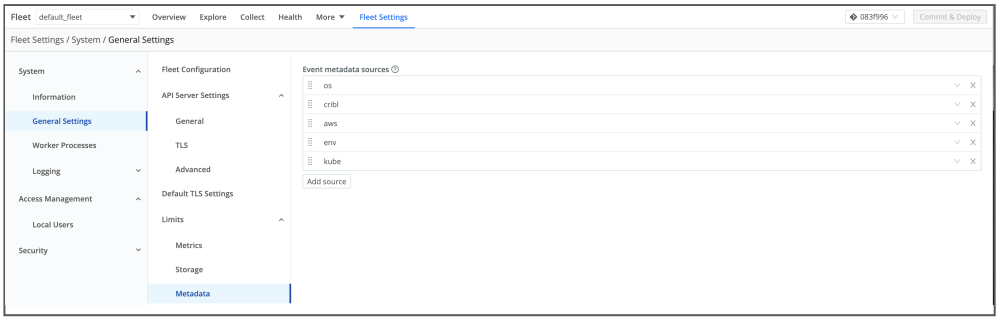
- `CRIBL_SERVICEACCOUNT_PATH` – by default, set to: `/var/run/secrets/kubernetes.io/serviceaccount`.
- `CRIBL_K8S_POD` – the name of the Kubernetes Pod in which Cribl Edge is deployed.
- `KUBECONFIG` – set this for local testing.
- `CRIBL_K8S_FOOTGUN`– set to `true` to enable resource-intensive, potentially risky modes of the Kubernetes Metrics and Kubernetes Logs Sources.
- `CRIBL_K8S_TLS_REJECT_UNAUTHORIZED` – set to `0` to disable certification validation when connecting to the Kubernetes APIs. When you disable this environment variable, all Kubernetes features (including Metadata, Metrics, Logs, and AppScope metadata) will tolerate invalid TLS certificates (i.e., expired, self-signed, etc.) when connecting to the Kubernetes APIs.

For details on other environment variables you can set for your Kubernetes deployment, see Environment Variables.

Metadata

You can customize the type of metadata collected at **Fleet Settings** > **Limits** > **Metadata**.

In Edge mode, all the Event metadata Sources are enabled by default. Leave the `kube` metadata set to allowed (by default).



Set limits

The `kube` metadata (`__metadata.kube property`) reports details on a Kubernetes environment, including the node, Pod, and container.

For the `__metadata.kube property` to report details on a Kubernetes environment, some level of Kube API access is required if you want to use this metadata property.

`KUBE_K8S_POD`, an environment variable in the Cribl Helm Chart, is set by default to the name of the Pod where Cribl Edge is running by default. This adds the `__metadata.kube` property with information to report on the Nodes and Pods.

Configuring the Kubernetes Sources

For the Kubernetes Sources to work as designed, Cribl recommends deploying Cribl Edge as a Daemonset.

Kubernetes Logs Source

The Kubernetes Logs Source connects to the Kubernetes API and loads the lists of Pods on the node, on a configurable Polling interval.

Set the polling interval based on the expected start and stop intervals of pods in your environment. Higher volume log sources may require a faster polling interval to avoid rotating logs faster than we can collect them.

For example, if the container emits `1 MBps` and rotates logs at `10 MB`, the polling interval must be less than `10s`, to prevent hitting the `10 MB` rotation threshold twice in 10 seconds.

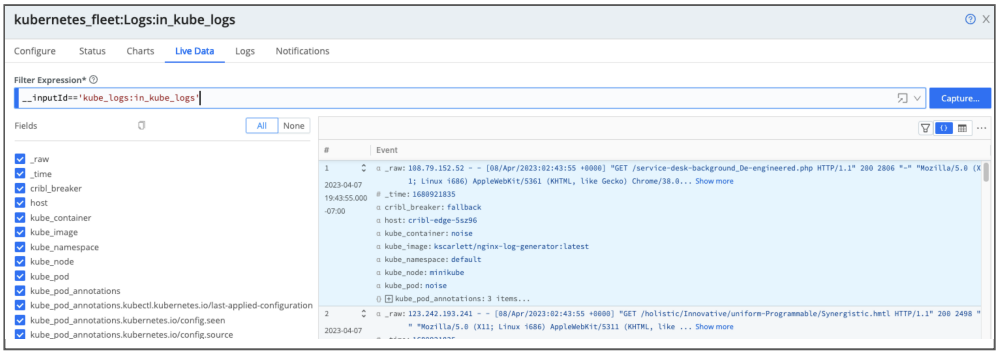
The Source then runs the Pods through the Filter Rules to determine which ones to report on. For example:

- Ignores Pods in the `kube-*` namespace - !metadata.namespace.startsWith('kube-')
- Ignore all DaemonSets – metadata.ownerReferences[0].kind != 'DaemonSet'

- Ignore Pods with specific Container names – spec.containers[0].name != 'edge'

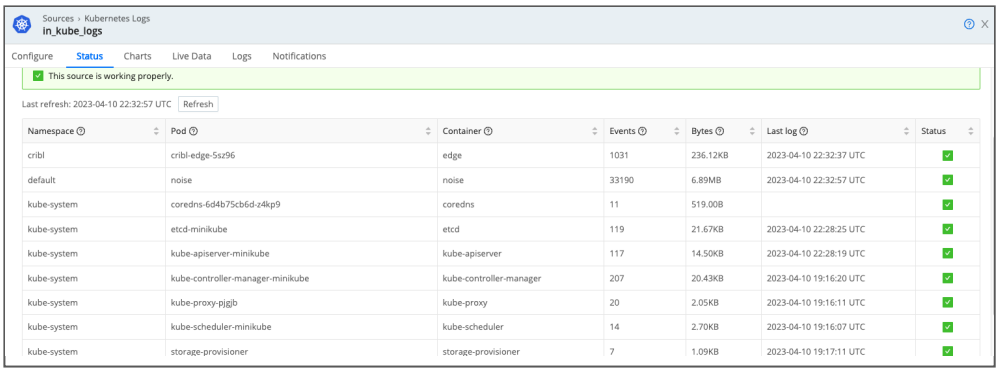
Configure the Kubernetes Logs Source to collect the Standard output (`stdout`) and Standard error (`stderr`) log streams from each Pod deployed inside the cluster.

Optionally, you can filter and enrich incoming logs.



Live Data: Kubernetes Logs Source

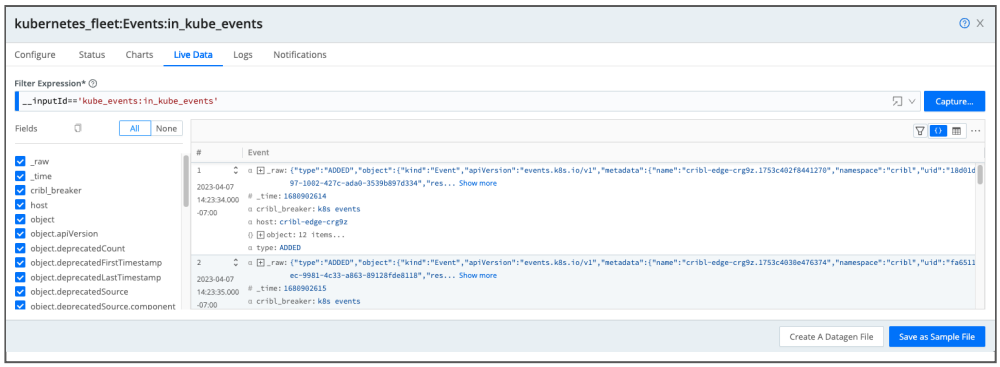
Below the **Capture** button, click the ellipsis (three dots or **...**) and then enable the **Show Internal Fields** toggle to explore the `__metadata` fields forwarded from the Cribl Edge Nodes.These metadata fields show information about the Pod, Node, and Cribl environment where they were collected.



Status: Kubernetes Logs Source

Kubernetes Events Source

Configure the Kubernetes Events Source to monitor state changes (such as Pod starts and jobs that fail to start) occurring on the Kubernetes cluster regarding workloads.



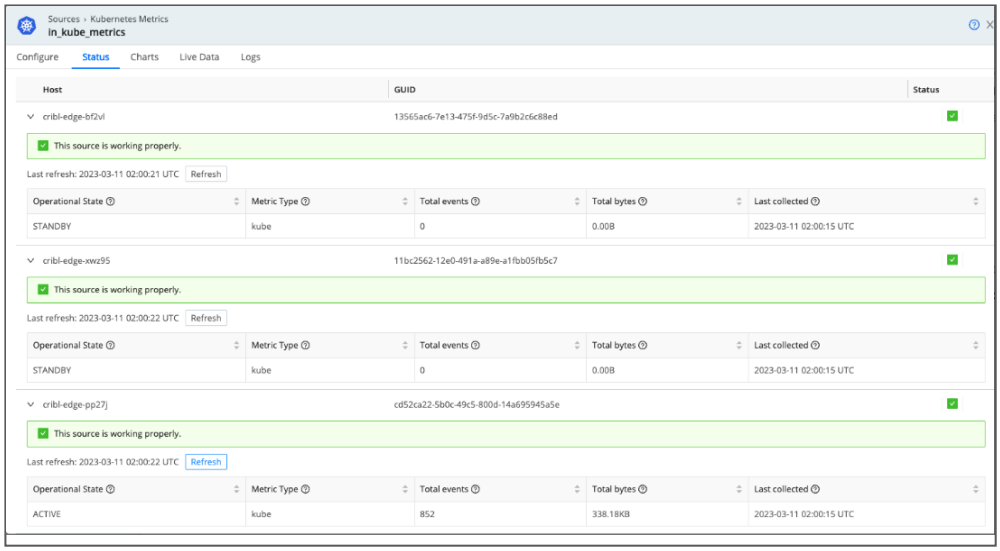
Live Data: Kubernetes Events Source

Kubernetes Metrics Source

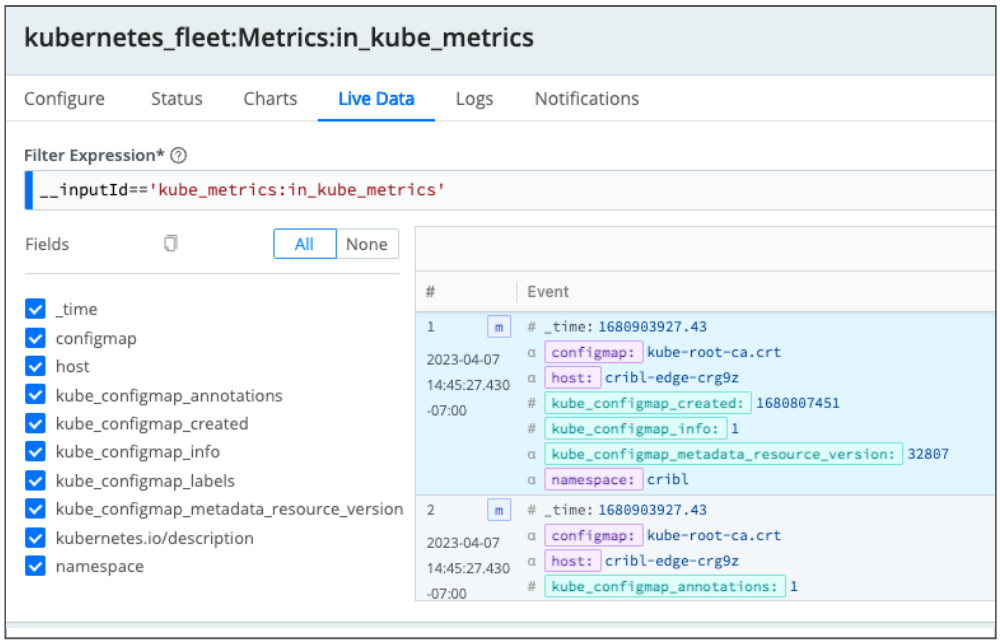
Configure the Kubernetes Metrics Source to generate metrics periodically based on the status and configuration of a Kubernetes cluster and its Nodes, Pods, and containers.

To avoid collecting the same metrics from multiple nodes, the Kubernetes Metrics Source uses an election method to collect metrics from only one node. The election happens every 5 minutes on all of the nodes, and the oldest node in the cluster wins.

To check the Source's operational state, go to **Status** and expand the host details. The **Operational State** column shows either an **Active** or **Standby** state. **Active** indicates the Source is running or won the election. **Standby** means it's waiting to be re-elected and not currently running.



Operational State



Live Data: Kubernetes Metrics Source

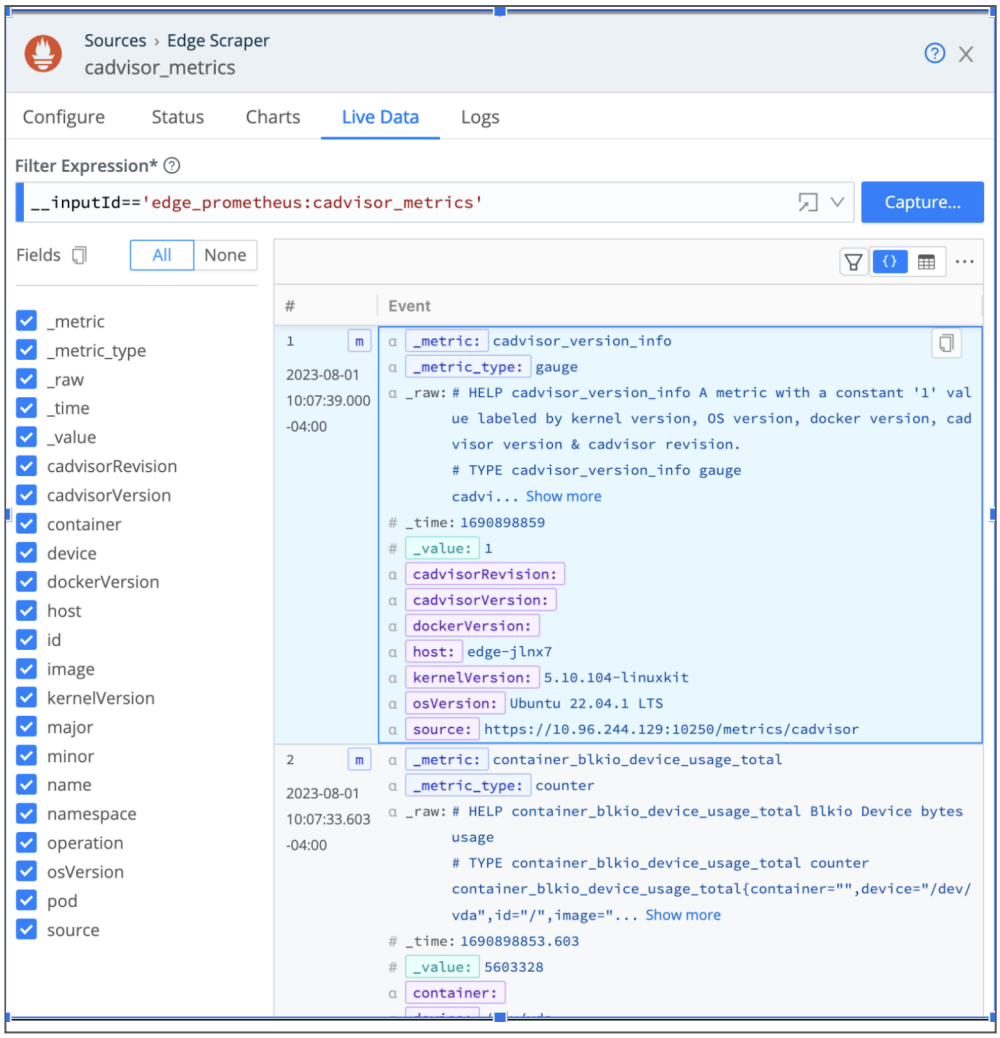
Prometheus Edge Scraper

You can also configure the Prometheus Edge Scraper to collect metrics from Pods and Nodes (as well as other Prometheus targets).

To collect Kubernetes metrics, Cribl Edge uses the `/metrics` and `/metrics/cadvisor` endpoints on the kubelet. This information gives us a really good idea of what's going on on each Node and Pod. In addition, this Source supports Disk Spooling and allows Cribl Search to query the metrics.

Using the Prometheus Edge Scraper, you can discover and collect metrics in Kubernetes Clusters dynamically. There are two Kubernetes-specific modes of operation:

- **Node**: When configured to scrape Nodes, each Cribl Edge Node (Pod) will scrape from the configured port on every deployed Node in the DaemonSet. Use this to scrape endpoints on the Kubelet and other node-local interfaces.
- **Pods**: When configured for Pod scraping, a single Edge Node will be elected to scrape the metrics from the discovered endpoints.



Live Data: Prometheus Edge Scraper

Configuring Your Destinations

For this example, we're going to use Prometheus and Loki for receiving metrics, logs, and events from Cribl Edge.

Loki stores log contents unindexed, relying on labels and timestamps that are indexed to form event streams. The method is similar to how data would be stored in a bucket or index on other systems.

Configure the Loki Destination to send log events to Grafana's Loki log aggregation system.

Similarly, Prometheus uses names, labels (dimensions), and numerical values. You can then create dashboards to visualize the change in data points over time using these time series.

Configure the Prometheus Destination to forward the Kubernetes Metrics to a Grafana dashboard (for our purposes).

Configuring the Pipelines

We will create two Pipelines to process the data as it routes to the configured Destinations.

Connecting Logs and Events

As previously mentioned, Loki stores logs based on labels. This is a great way to organize your stored data so that retrieval is easier when looking for specific log entries. Configure a Pipeline with an Eval Function to add the labels to the `__labels` field.

By default, no labels are defined for the Loki Destination. Because Loki requires at least one label, Cribl will automatically add a default `label` source with the value `cribl_${__inputId}`.

Once you create a Pipeline, add an Eval Function with the following config in the **Evaluate Fields** section:

Name	Value Expression
`__labels`	`{}`
`__labels.namespace`	`kube_namespace`
`__labels.pod`	`kube_pod`
`__labels.container`	`kube_container`

Next, go to **Manage** > **Collect** to attach the Kubernetes Logs Source to the Loki Destination, using the configured Pipeline as the connection.

You can connect the Kubernetes Events Source directly to your configured Loki Destination with the default `Passthru` Pipeline connection.

Connecting Metrics

You can modify the default `prometheus_metrics1` Pipeline to change a few metadata fields and use it to connect to the Prometheus Destination.

In the `prometheus_metrics1` Pipeline, expand the Eval Function with the following config in the **Evaluate Fields** section:

Name	Value Expression
`job`	job || 'cribl'
`instance`	__metadata.kube.node.metadata.name || host

Name	Value Expression

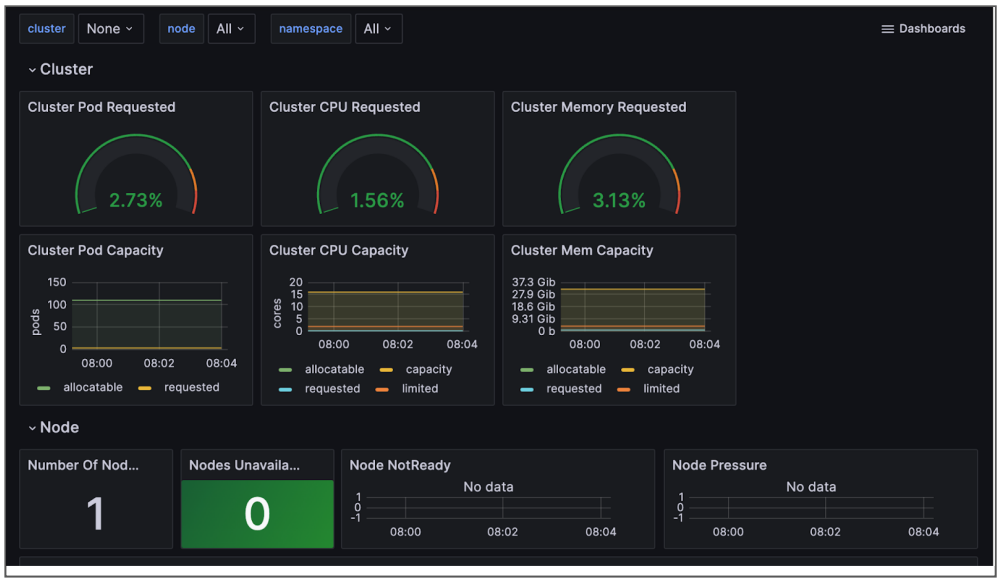
Next, go to **Manage** > **Collect** to attach the Kubernetes Metrics Source (or Prometheus Edge Scraper) to the Prometheus Destination, using the `prometheus_metrics1` Pipeline as the connection.

Visualize the Data in Grafana

Navigate to your Grafana Cloud instance, and select Explore on the left panel to review the Node metrics. Selecting metrics and running the query will produce a chart and table of metric events that can be added to new or existing dashboards for analysis.

If you are missing some expected metrics, you might need to adjust the variables in your Grafana Cloud instance. For details, see Grafana's Variables topic.

Use Grafana's Filter option to browse the dashboards for your Prometheus data and Loki Logs. Below is a sample dashboard.



Grafana Dashboard

5.4. Windows Observability Using Cribl Edge

Cribl Edge offers holistic support for Windows observability, allowing comprehensive data collection across your Windows servers to provide reliable monitoring and analysis. You can run Cribl Edge on your Windows servers (2016, 2019, or 2022) to collect observability data (both metrics and logs) and send them to your desired destinations. By combining Cribl Edge with your favorite system of analysis, or data visualization tool, you can identify and investigate issues faster and monitor applications more effectively.

In this guide, we will walk you through the following:

- Deploy Cribl Edge on Windows by using the bootstrap script to add an Edge Node. We will also cover how to create a Windows-specific Fleet to connect to your Cribl Leader and Mapping Rulesets to switch Fleets.
- Explore the Windows Node and get a snapshot of the host system's current state.
- Locate and collect custom Windows event logs and send them to your desired destination.
- Collect Windows metrics and send them to your visualization tool of choice (or any supported destination).
- Collect logs using the File Monitor Source.

As a bonus, we'll walk you through how to use Cribl's Exec Source to execute Powershell commands and process them in Cribl.

Deploy Cribl Edge on Windows

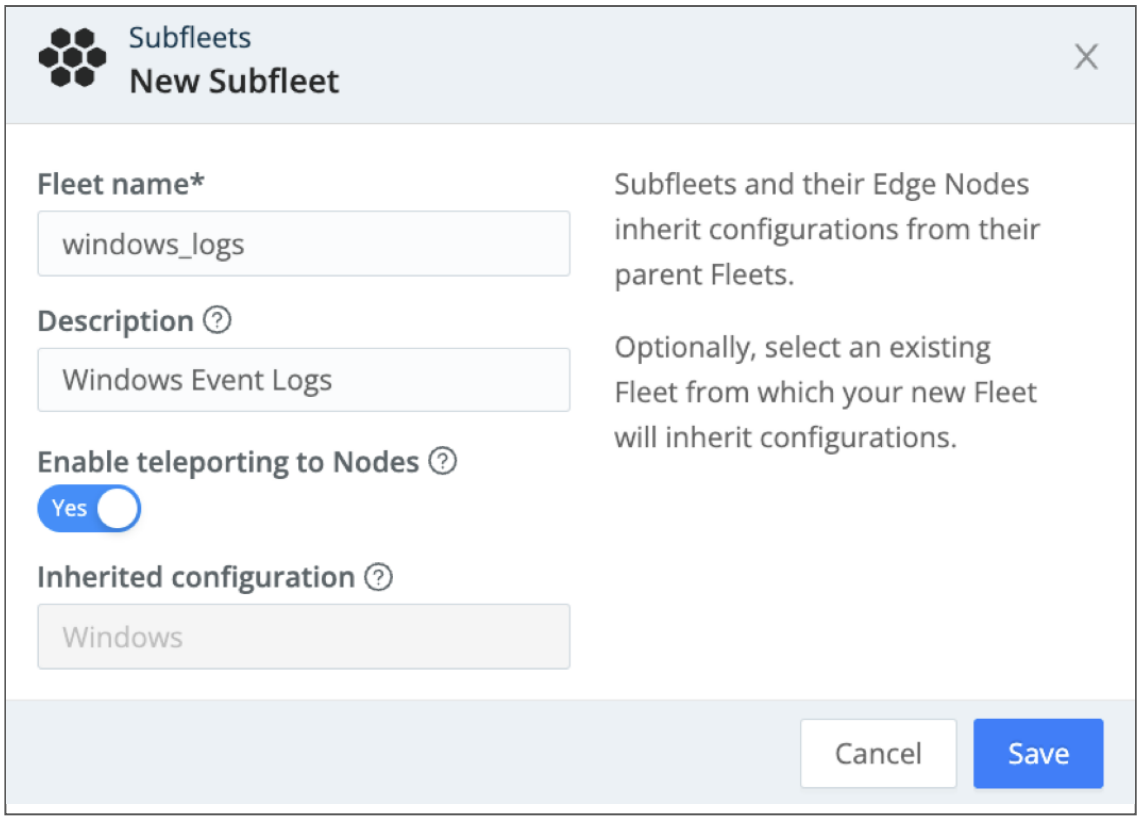
This section covers creating a Windows-specific Fleet, adding Windows Nodes via bootstrap scripts, and assigning Nodes to Fleets or Subfleets via mapping rules.

Create a Windows-Specific Fleet

Before we deploy the Edge Node, let's configure a Fleet. Fleets and their corresponding Subfleets allow you to author and manage configuration settings for a particular set of Edge Nodes. We recommend separating Linux and Windows machines into different Fleets per OS.

The first step is to configure a Fleet and name it: `Windows`.

For our mapping illustration later, let's also create a Subfleet in your `Windows` Fleet. In our example, we named it `windows_logs`.



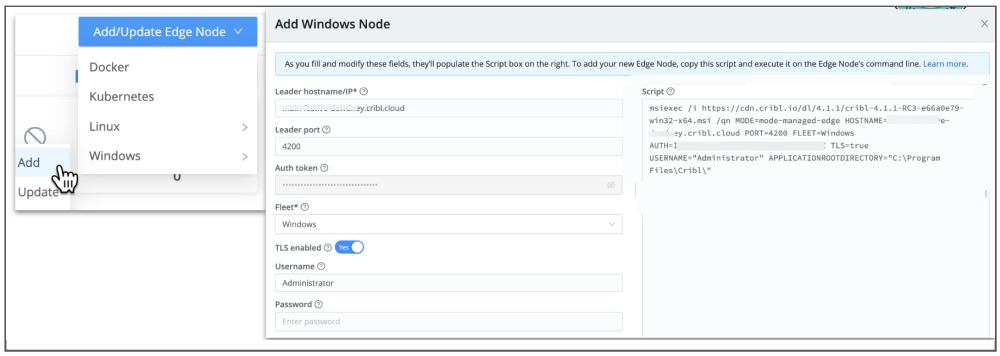
Create a Subfleet

We use the name `Windows` for demonstration purposes and simplicity. Since Fleet names must be unique across your deployment, you should select a name that makes sense in practice. For better practices on designing and naming your Fleets, see the Fleet Hierarchy and Design doc.

Use the Bootstrap Script to Add an Edge Node

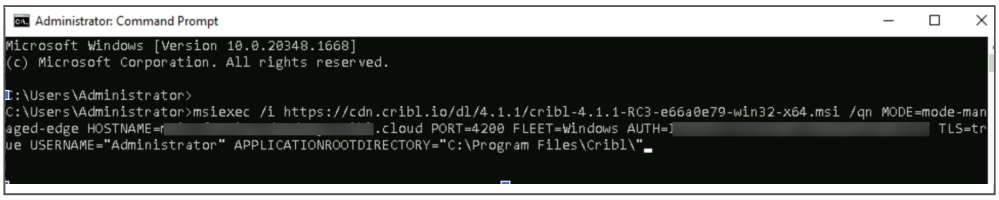
Next, bootstrap a new Windows Edge Node on your Fleet.

Below is a composite screenshot of adding an Edge Node on Windows.



Generating a script to bootstrap an Edge Node on Windows

To add the Edge Node, paste the script into your Windows command prompt and execute it. You must run the command line as an administrator for this to work. For details, see Start Command Prompt as an Administrator.



Pasting the script to the Command Prompt

You can edit the `instance.yml` files, to change the following:

- Reassign the Edge Node to a different Fleet.
- Enable TLS after installation for prior versions of Cribl Edge. In Cribl Edge 4.1 and later, this is included in the bootstrap script when you deploy from a Cribl.Cloud Leader. For details, see Enabling TLS After Installation.

For details, see Using YAML Config File to configure an Edge Node.

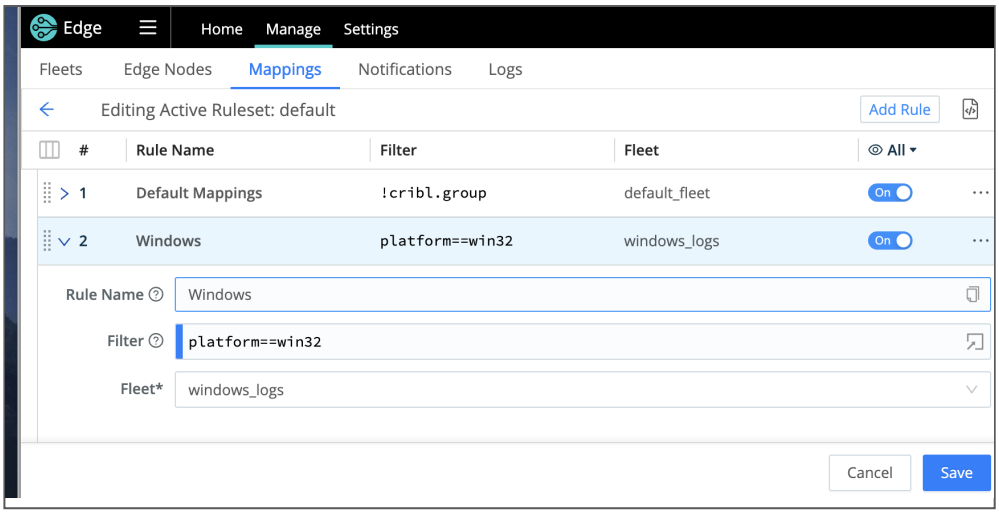
Keep in mind, there were two more changes related to deployment as of Cribl Edge 4.1:

- The 4.1 (and later) MSI installer uses the `CRIBL_VOLUME_DIR`, so logs, configs, state, etc. are in `ProgramData\Cribl` instead of `Program Files\Cribl`. The `instance.yml` is located in ProgramData\Cribl too.
- The 4.1 (and later) MSI installer's command-line syntax changed, so if you simply change the MSI file it uses and reuse the earlier installation command line, you will get default settings, i.e., standalone Edge. The current syntax can be found in the Add Node UI.

Use a Mapping Ruleset to Switch Fleets

In this step, we will show you how to create a Mapping Ruleset to switch the Windows Edge Node to the Subfleet `windows_logs` you created earlier.

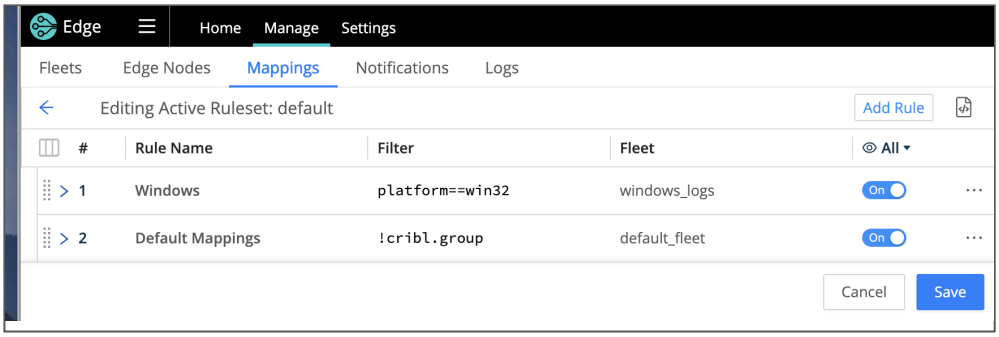
First, we'll add a new rule in the existing `default` ruleset.



Map to the SubFleet

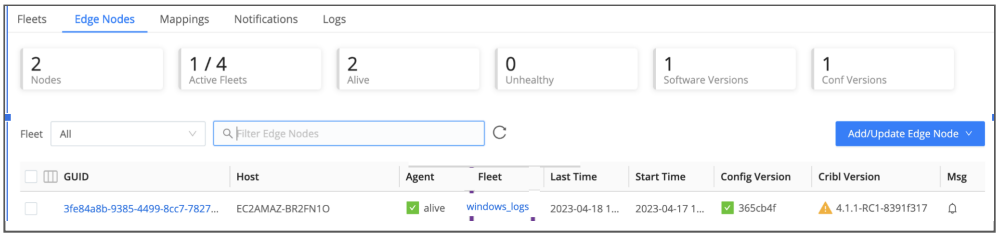
Only one Mapping Ruleset can be active at a time, although a ruleset can contain multiple rules. It's important to keep your Fleet assignments within one ruleset. This example uses the `default` ruleset.

Make sure you reorder the rules, so the new rule is first.



Reorder the Rulesets

Your Windows Edge Node will now be assigned to the new SubFleet.



View Edge Nodes

For details on mapping, see Mapping Edge Nodes to Fleets.

Explore Your Windows Node

When you first log in to Cribl Edge on Windows (single-instance or managed node), you'll land on the **Home** tab where you can explore the metrics and log data that the Node has auto-discovered and manually discover and explore other data of interest. For details, see Exploring Cribl Edge on Windows.

From the Fleet's **Explore** menu, click the **System State** tab to see snapshots of the host system's current state in configurable time intervals. For details on what's displayed, see System State.

You can specify which collectors are enabled and the polling interval, by configuring the System State Source.

Configure the Windows Event Logs Source on Cribl Edge

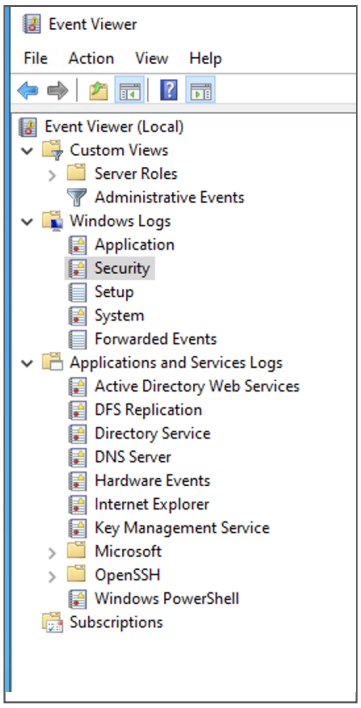
Locate Windows Event Logs in the Server

In order to configure the Windows Event Logs Source on Cribl Edge, you need to figure out which event logs you want to collect from your Windows Server.

Event logs in Windows Servers are classified into the following broad categories:

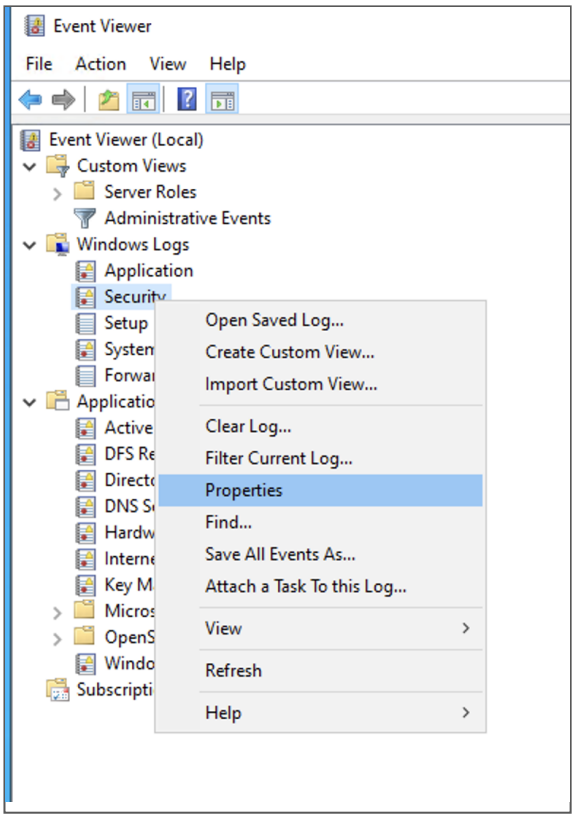
- **System**: Events related to the system and its components. Failure to load the boot-start driver is an example of a system-level event.
- **Applications and Services**: Events related to software or an application hosted on a Windows computer get logged under the application event log. For example, if a user encounters a problem in loading the app, it will be logged.
- **Security**: Events related to the safety of the system. The event gets recorded via the Windows auditing process. Examples include failed authentication and valid logins.
- **Setup**: Events occur during the installation of the Windows operating system. On domain controllers, this log will also record events related to Active Directory.
- **Forwarded Events**: Contains event logs forwarded from other computers in the same network.

To find out what other event logs might be of interest, go to your Windows Server Event Viewer.



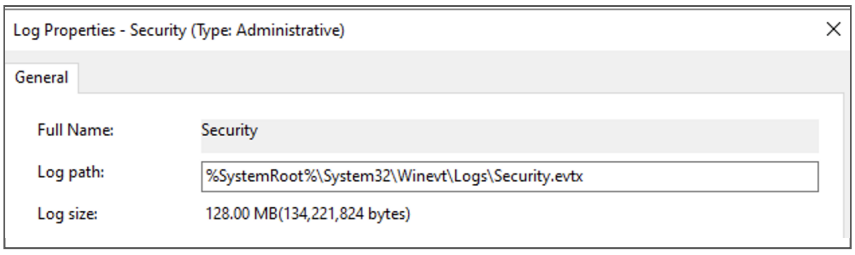
Event Viewer

To get the proper name of the event log, right-click and select **Properties**.



Log properties

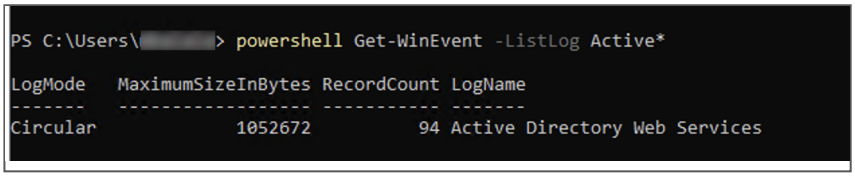
In the Windows Event Log Source's configuration modal, copy the value in **Full Name** and paste it into **Event Logs**.



Log properties: Full name from the UI

Alternatively, you can execute the following PowerShell command to list a particular log's name. In this example, we are using the * as a shortcut.

```
powershell Get-WinEvent -ListLog Hardware*
```

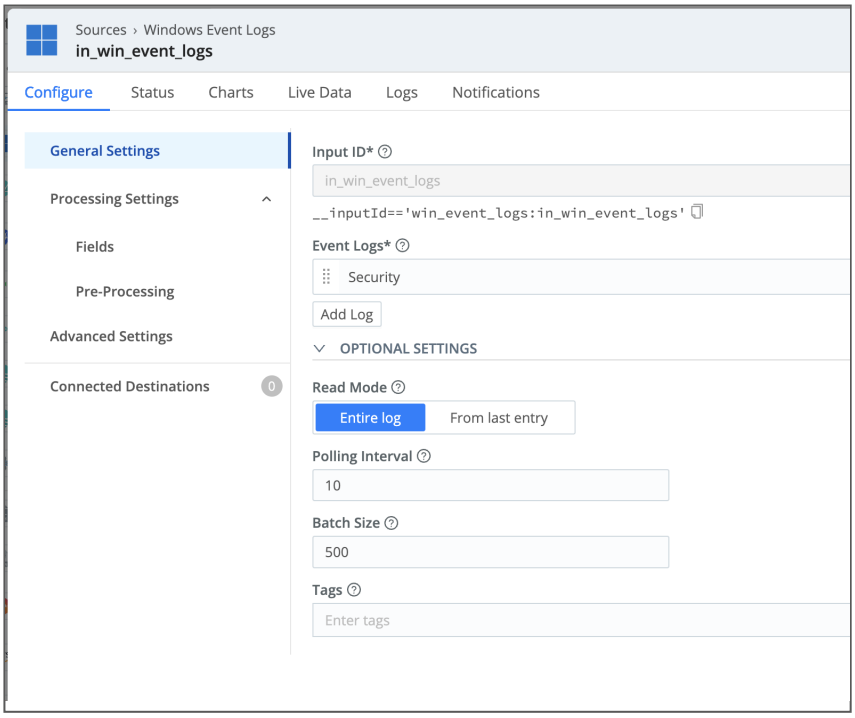


Log Name from PowerShell

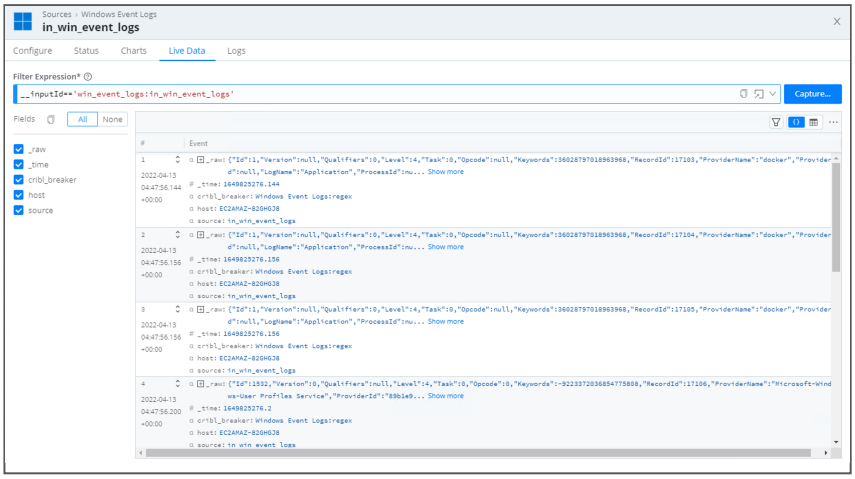
From the powershell output, the **LogName** field is what you would specify in the Cribl UI. In the above screenshot, for example, you would enter `Active Directory Web Services` in the Cribl UI.

Configure the Windows Event Logs Source

Once you have the accurate names of the event logs you want to collect, you can configure and enable the Windows Event Logs Source in Cribl Edge.



When done, **Commit** and **Deploy** your changes. Before moving on to the next step, check the **Live Data** tab to make sure the logs are generated.



Live Data

You can now send the event logs to your visualization tool of choice or any of Cribl Edge's supported Destinations.

Configure the Windows Metrics Source on Cribl Edge

In Cribl Edge, start by configuring and enabling the Windows Metrics Source.

The key requirement here is to decide on the level of granularity for the metrics you want to collect. In the **Windows Metrics** Source, go to the **Host Metrics** tab, and select the following options for each type:

- **CPU**: Set to **Custom**, enable **Per CPU** metrics.
- **Network**: Set to **Custom**, enable **Per Interface** metrics.
- **Disk**: Set to **Custom**, enable **Per Volume** metrics.

On the **Pre-Processing** tab, select the `prometheus_metrics` Pipeline.

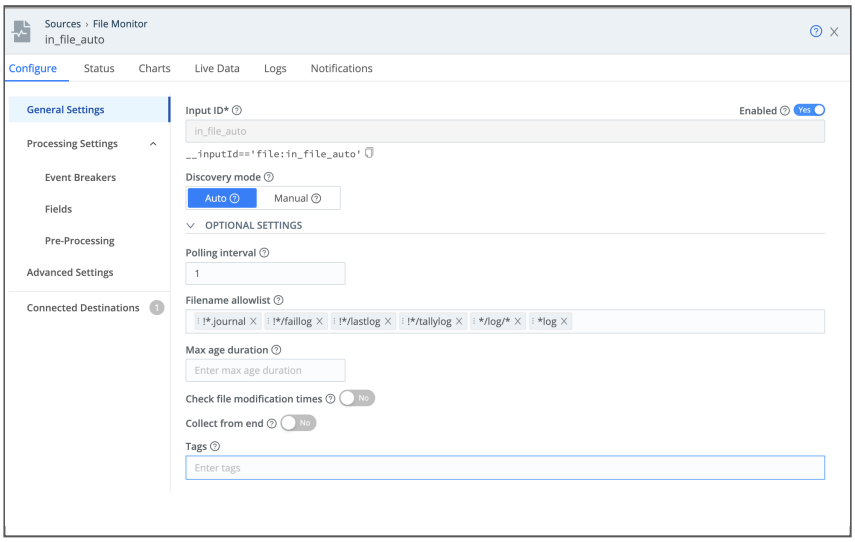
When done, **Commit** and **Deploy** your changes. Before moving on to the next step, check the **Live Data** tab to make sure the metrics are generated.

To see an example of how to connect to your Windows Metrics Source to a Grafana Cloud Destination and visualize the data in Grafana, see Monitoring your Infrastructure with Cribl Edge.

Configure the File Monitor Source on Cribl Edge

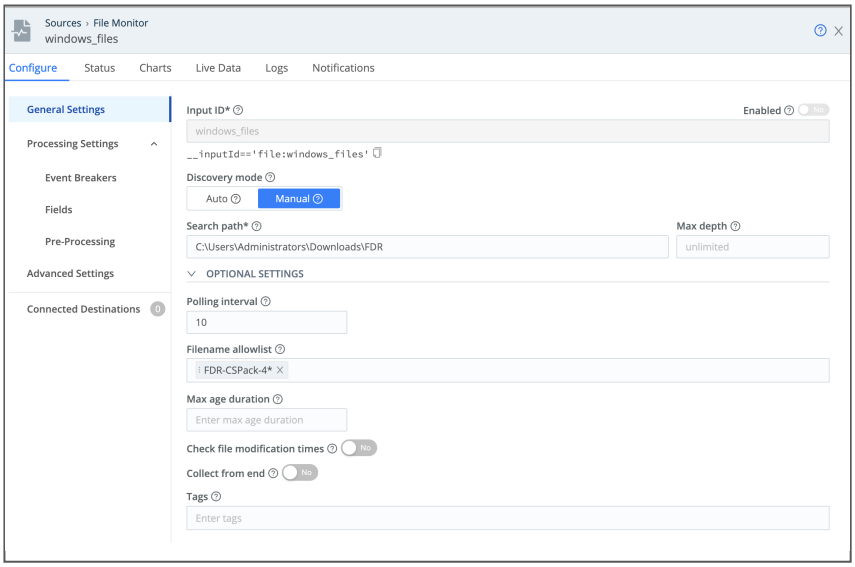
In Cribl Edge, start by configuring and enabling the File Monitor Source.

This configuration monitors all files ending with `.log`.



File Monitor in Auto mode

The **Manual** configuration option allows you to monitor a select file within a directory. When creating a new entry, specify the folder where the log files can be found. If the **Max depth** is empty, all subdirectories will be searched. To limit it to the referenced directory only, set **Max depth** to `1`. Update the **Filename allowlist** with the specific file to monitor; this field supports wildcards. For guidance on how to use allowlists without creating duplicates, see Using Allowlists Effectively.



File Monitor Configuration in Manual mode

Bonus Section: Executing PowerShell Commands

The Exec Source enables you to periodically execute a command and collect its `stdout` output. This is typically used in cases when Cribl Edge cannot accomplish collection with native Collectors or other Sources.

The Exec Source command is actually executed from a Windows Command line, not a PowerShell command line.

The execution of PowerShell is resource-intensive and should be used with caution. On limited resource systems (e.g., domain controllers) it is best to run Exec with PowerShell carefully as the Source uses the host's CPU.

The main thing to keep in mind when you configure Cribl's Exec Source is that what you enter in the **Command** field is slightly different from a PowerShell prompt command.

PowerShell Prompt Command	Cribl Exec Source configuration
`Get-Process`	`powershell Get-Process`
`# cd c:\mydir` `# .\myps-script.ps1`	`powershell c:\mydir\myps-script.ps1`

You can handle a PowerShell command with parameters as follows:

To process the PowerShell command for this script: `Get-CimInstance win32_process | select ProcessId,Name, commandline,HandleCount,WorkingSetSize,VirtualSize`

Enter the following into the Exec Source's **Command** field:

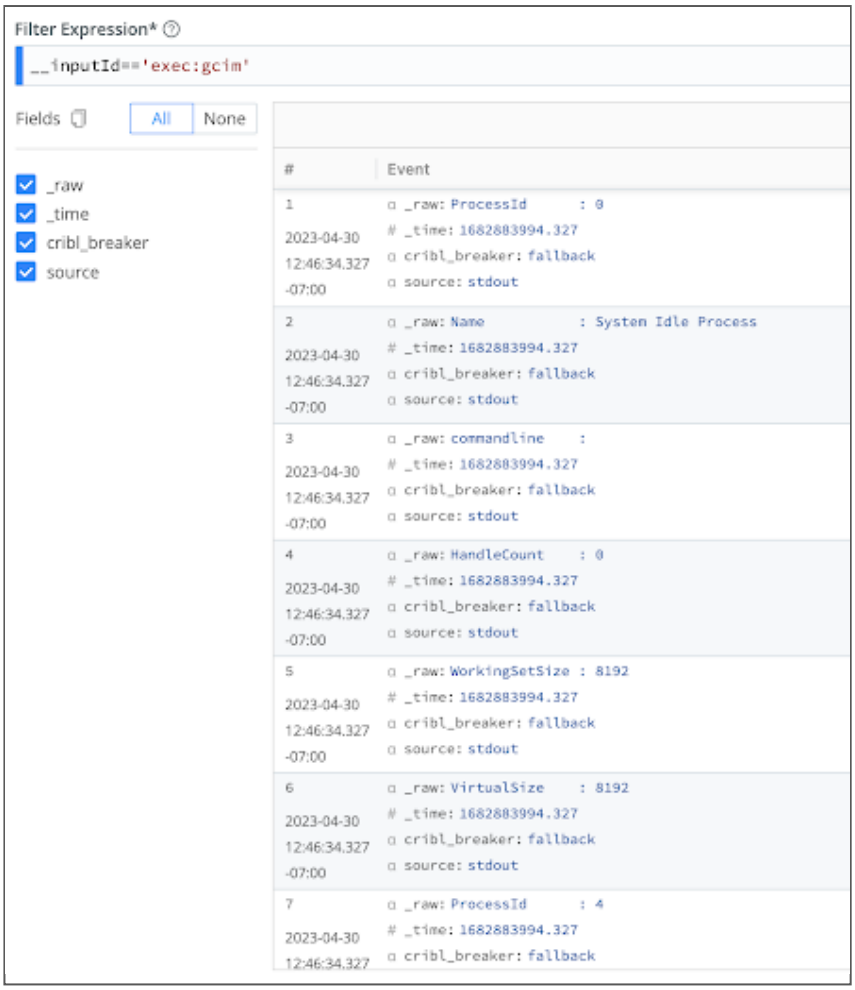
```
powershell -command "Get-CimInstance win32_process | select ProcessId, Name, commandline, HandleCount, WorkingSetSize,VirtualSize"
```

And for the newer version of the Powershell command, you can enter the following into the Exec Source's **Command** field:

```
powershell "Get-CimInstance win32_process | select ProcessId, Name, commandline, HandleCount, WorkingSetSize,VirtualSize"
```

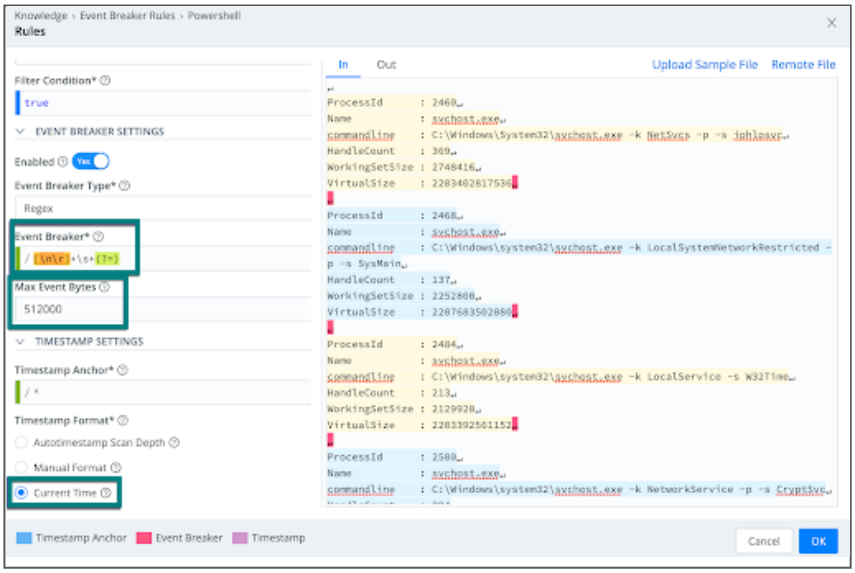
Exec Source's Command

The output of the command without using an Event Breaker:



Default Output

The output of the command with an Event Breaker:



Output with an Event Breaker

6. Workspaces

Workspaces offer a multi-tenancy capability, enabling you to create multiple isolated instances in your Cribl.Cloud Organization. This lets you strengthen your security, and fulfill compliance and isolation requirements.

Each Workspace offers a dedicated virtual public cloud (VPC) that acts as a separate environment within your Organization.

You can manage all your Workspaces in one interface that contains access controls for granting Members and Teams access to individual Workspaces.

Workspaces require a Cribl.Cloud Enterprise plan.

An example use case scenario for creating multiple Workspaces is setting up separate environments for different business units in an enterprise. Data management, security, development, and any other units can have their own federated Workspaces, with completely separate members and permissions lists. This way, they can work in isolation without risk of interference and with increased focus.

Opt-In to Workspaces

Access to configuring multiple Workspaces is available once you opt in to a new Cribl.Cloud UI:

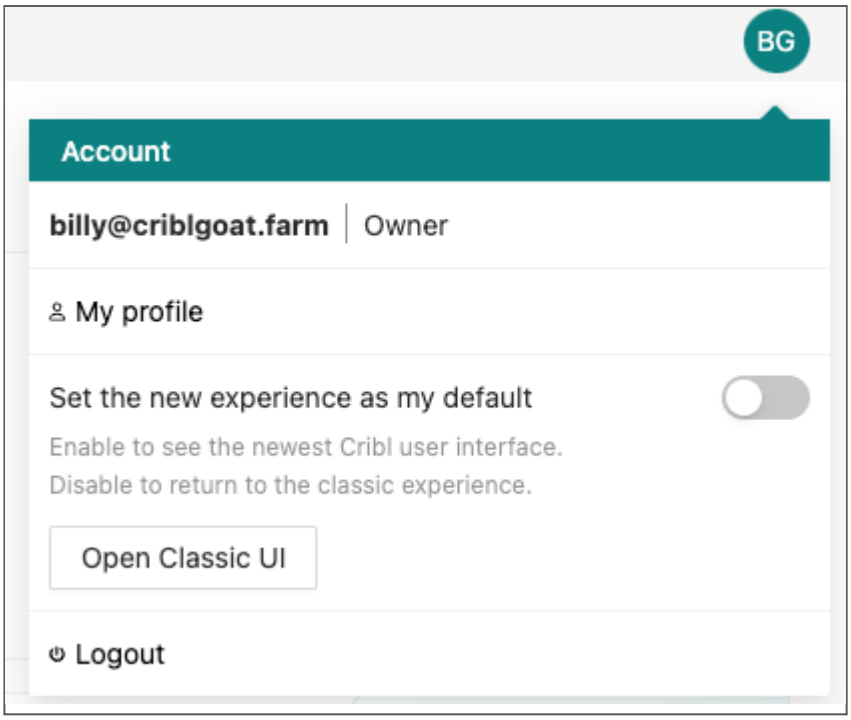
1. On the Cribl.Cloud top bar, click **Try the New Cribl.Cloud**.
2. In the modal, confirm with **Get Started**. You will move to a new user interface for managing Cribl.Cloud, including Workspaces.

Your Cribl.Cloud Organization will open with the new UI every time you re-enter it. You can change this behavior in the following way:

1. Open your user menu at the right side of the top menu.
2. Disable **Set the new experience as my default**.

If you want to return to the old UI:

1. Open your user menu at the right side of the top menu.
2. Disable **Set the new experience as my default**.
3. Click **Open Classic UI**.



You can return to the classic UI at any time

Only users with the **Organization Owner** or **Organization Admin** Member permission can opt in to the new user interface.

Limitations

You can create up to 5 Workspaces per Organization.

Workspaces require a Cribl.Cloud Enterprise plan.

6.1. Configuring Workspaces

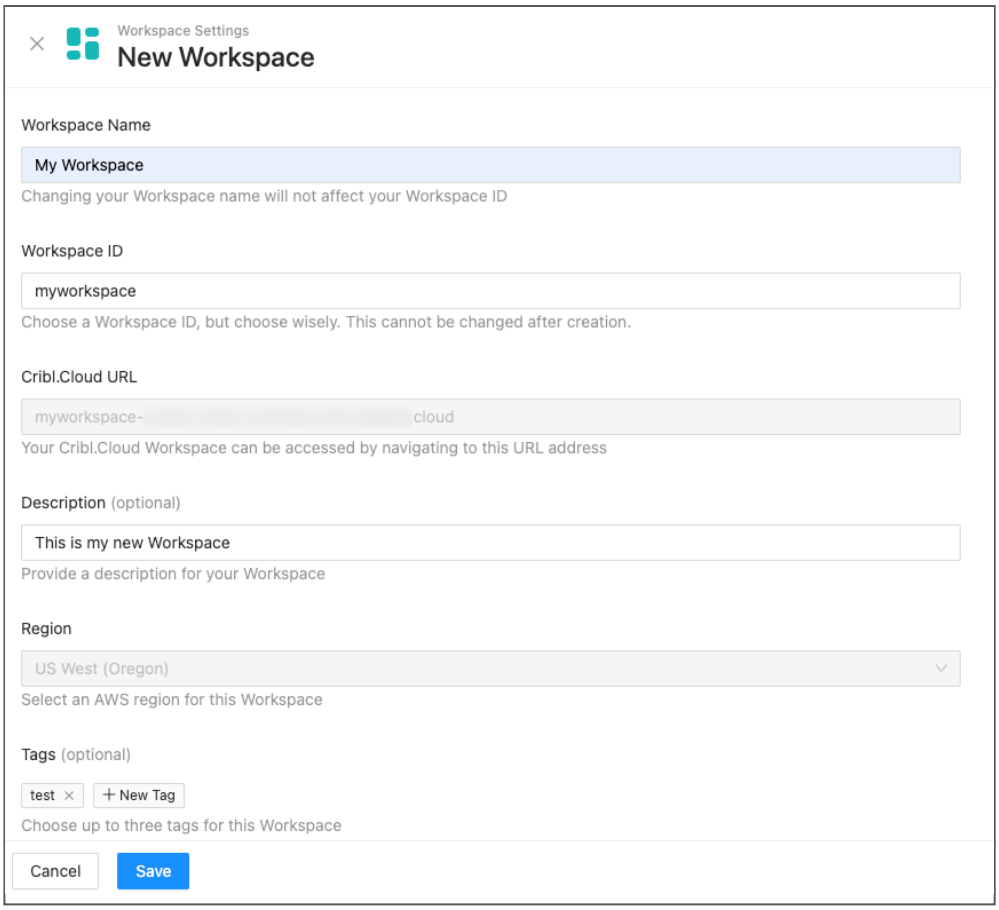
Add New Workspace

Before you try creating a new Workspace, make sure you have opted in to the new Cribl.Cloud UI.

To create a new Workspace:

1. From the top bar of your Organization's front page, click your current Workspace name. This opens a modal with a list of all Workspaces you have access to.
2. Select **Add Workspace**.
3. Fill the form with the following information:

Field	Description	Notes
Workspace Name	Human-readable name for the Workspace.	
Workspace ID	Internal ID for the Workspace. This ID will be part of the Workspace's Cribl.Cloud URL.	**Cannot be changed** once the Workspace is created.
Description	Additional Workspace description.	Optional
Region	Set automatically to correspond to the Organization region.	
Tags	Up to 3 tags you can use to filter Workspaces.	Optional



Creating a new Workspace

1. Confirm with **Save**.

The Workspace can take up to several minutes to spin up. While the process continues, you can navigate away and keep working with your Cribl.Cloud instance as usual.

Edit a Workspace

You can edit your Workspace at any time, by clicking the gear ( ) button in the top right corner. You can change the name, description, and tags of a Workspace. However, you can't change its identifier once it has been created.

Delete a Workspace

To delete a Workspace:

1. Click the gear ( ) button in the top right corner of the Workspace's screen.
2. Select **Delete Workspace**.

Only Members with the Organization Owner permission can delete Workspaces.

6.2. Members And Teams

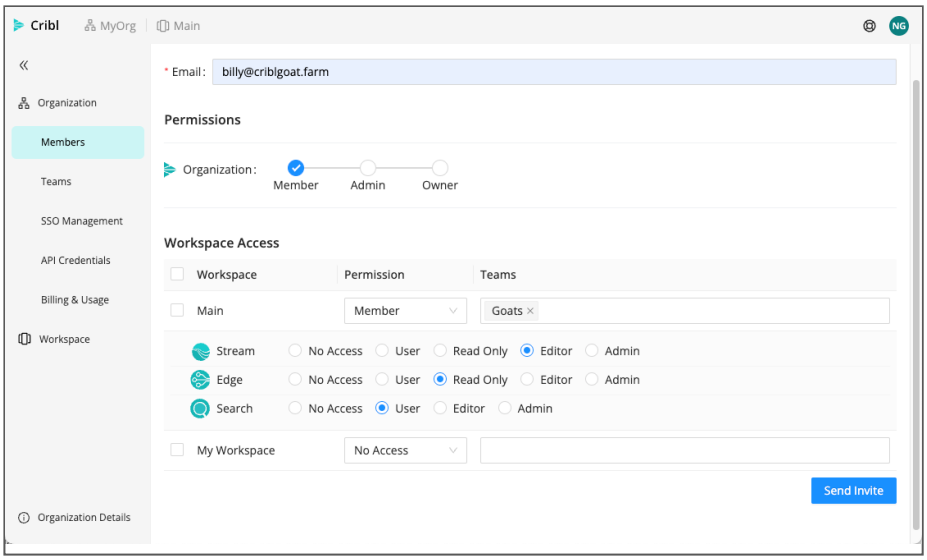
One of the main benefits of Workspaces is the ability to granularly control which users have access to which parts of the system. Access management is done though assigning permissions to individual Members and Teams within a Workspace.

If a Member has overlapping assigned permissions, the highest permission is given.

Members

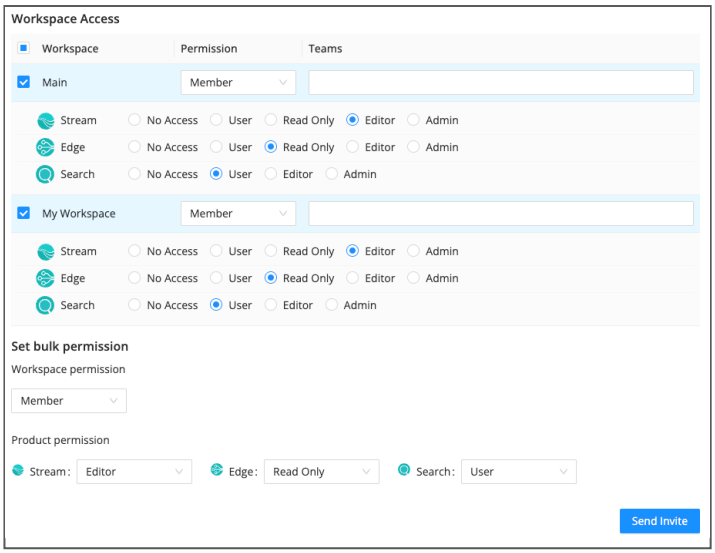
When inviting a new user to your Organization, configure permissions for Workspaces in the **Workspace Access** section:

1. Select a Member permission in the **Permission** dropdown next to the Workspace.
   When you choose the **Member** permission, configure access for each Cribl product separately:
2. Select **Teams** to assign to the Member.



Selecting product-level permissions and a Team for a new Member.

Alternatively, select the check box next to one or more Workspaces and choose a permission under **Set bulk permission**. This permission will then be granted to the user for all selected Workspaces.



Setting bulk permissions for two Workspaces.

Teams

Teams are groups of Members who share the same Workspace and product-level permissions. This allows you to efficiently manage access control by assigning permissions to the Team rather than configuring each Member individually.

To configure a Team:

1. On the left-side menu, expand **Organization** and then select **Teams**.
2. Select **Create Team** and the configuration modal is displayed.
3. Provide a meaningful **Name** and optionally a **Description**.
4. Select the desired Workspace and access to assign to the Team. **Members** allow granular product-level permissions.
5. In **Team Members**, select the Members you want to add, then **Add** them. A Member must already exist to be added to a Team.
6. The table is updated to show the Members you've just added.
7. When finished, **Save** your new Team.

The main Teams page displays the Teams configured on your Workspace. You can **Edit** or delete them as needed from the **Actions** column.

7. Administering

7.1. Upgrading Overview

Learn about the ways you can upgrade Cribl Edge, as well as general and version-specific upgrade considerations.

Ways to Upgrade Cribl Edge

You can upgrade Cribl Edge in two ways:

- Upgrade from the UI.
- Upgrading Manually allows you to upgrade by uncompressing a new version of Edge on top of the old one.

General Upgrade Considerations

Here are some considerations to review before you upgrade.

- Upgrading Edge Nodes from the Leader requires a Standard or Enterprise license.
- Cribl Edge does **not** support direct upgrades from any pre-GA version (such as a Cribl-provided test candidate) to a GA version. To get the GA version running, you must perform a new install.
- Before upgrading your Leader to v.4.0 and later, see Persisting Socket Connections to prepare the host to keep communications open from Edge Nodes.
- Cribl Edge 4.1 and later encrypts TLS certificate private key files when you add or modify them. See instructions just below for backing up your keys from earlier versions.

Leader and Edge Nodes Compatibility

Leaders on v.4.2.x are compatible with Edge Nodes on v.4.1.2, v.4.1.3, and later. Due to a security update, Edge Nodes running on v.4.0.4 cannot receive configurations from v.4.2.x Leaders.

Upgrading to v.3.5.4

If you're planning to upgrade to v.3.5.4, note that any Edge Nodes must be at the same version as the Leader for v.3.5.4. Leaders running v.3.5.4 and later should test whether Edge Nodes are running a compatible

version before deploying configs that could break Edge Nodes' data flow. The Leader will prompt you to upgrade these Nodes as needed.

To prevent Edge Nodes from failing on incompatible configurations, we introduced three new behaviors announced in Cribl Edge 3.5.4 release notes. These will make data throughput more resilient, but might enforce more-frequent (automated or explicit) upgrades to Nodes' Edge versions:

- The Edge Leader will block configuration deployments to Edge Nodes running an earlier Edge version that's incompatible with the new configs. The Leader will prompt you to upgrade these Nodes to a compatible version.
- If Edge Nodes detect a Source or Destination type for which they have no supporting configuration, they will ignore (skip initializing) that integration, rather than failing.
- For unsupported Destination types, Nodes will create a temporary placeholder Destination, but will exert Blocking backpressure until the configuration is updated. You will see a warning of the form: `Skipping configuring Destination...due to unrecognized type.`

Safeguarding Unencrypted Private Keys for Rollback

Cribl Edge 4.1 and later encrypt TLS certificate private key files when you add or modify them.

Before upgrading from a pre-4.1 version, make a backup copy of all unencrypted TLS certificate private key files. Having access to the unencrypted files is essential if you later find that you need to roll back to your previous version.

To safeguard your unencrypted private keys, make a full backup of all Cribl config files. Files in the `auth/certs` directory are particularly important, such as those in:

- `groups/default/local/cribl/auth/certs/`
- `groups/<groupname>/local/cribl/auth/certs/`
- `cribl/local/cribl/auth/certs/`
- Etc.

Take appropriate precautions to prevent unauthorized access to these unencrypted private key files. If you need to roll back to a pre-4.1 version, see Restoring Unencrypted Private Keys.

Keep Reading

- Upgrading via UI

- Upgrading Manually
- Troubleshooting Cribl Edge Upgrades

7.1.1. Upgrading Overview

Learn about the ways you can upgrade Cribl Edge, as well as general and version-specific upgrade considerations.

Ways to Upgrade Cribl Edge

You can upgrade Cribl Edge in two ways:

- Upgrade from the UI.
- Upgrading Manually allows you to upgrade by uncompressing a new version of Edge on top of the old one.

General Upgrade Considerations

Here are some considerations to review before you upgrade.

- Upgrading Edge Nodes from the Leader requires a Standard or Enterprise license.
- Cribl Edge does **not** support direct upgrades from any pre-GA version (such as a Cribl-provided test candidate) to a GA version. To get the GA version running, you must perform a new install.
- Before upgrading your Leader to v.4.0 and later, see Persisting Socket Connections to prepare the host to keep communications open from Edge Nodes.
- Cribl Edge 4.1 and later encrypts TLS certificate private key files when you add or modify them. See instructions just below for backing up your keys from earlier versions.

Leader and Edge Nodes Compatibility

Leaders on v.4.2.x are compatible with Edge Nodes on v.4.1.2, v.4.1.3, and later. Due to a security update, Edge Nodes running on v.4.0.4 cannot receive configurations from v.4.2.x Leaders.

Upgrading to v.3.5.4

If you're planning to upgrade to v.3.5.4, note that any Edge Nodes must be at the same version as the Leader for v.3.5.4. Leaders running v.3.5.4 and later should test whether Edge Nodes are running a compatible version before deploying configs that could break Edge Nodes' data flow. The Leader will prompt you to upgrade these Nodes as needed.

To prevent Edge Nodes from failing on incompatible configurations, we introduced three new behaviors announced in Cribl Edge 3.5.4 release notes. These will make data throughput more resilient, but might enforce more-frequent (automated or explicit) upgrades to Nodes' Edge versions:

- The Edge Leader will block configuration deployments to Edge Nodes running an earlier Edge version that's incompatible with the new configs. The Leader will prompt you to upgrade these Nodes to a compatible version.
- If Edge Nodes detect a Source or Destination type for which they have no supporting configuration, they will ignore (skip initializing) that integration, rather than failing.
- For unsupported Destination types, Nodes will create a temporary placeholder Destination, but will exert Blocking backpressure until the configuration is updated. You will see a warning of the form: `Skipping configuring Destination...due to unrecognized type.`

Safeguarding Unencrypted Private Keys for Rollback

Cribl Edge 4.1 and later encrypt TLS certificate private key files when you add or modify them.

Before upgrading from a pre-4.1 version, make a backup copy of all unencrypted TLS certificate private key files. Having access to the unencrypted files is essential if you later find that you need to roll back to your previous version.

To safeguard your unencrypted private keys, make a full backup of all Cribl config files. Files in the `auth/certs` directory are particularly important, such as those in:

- `groups/default/local/cribl/auth/certs/`
- `groups/<groupname>/local/cribl/auth/certs/`
- `cribl/local/cribl/auth/certs/`
- Etc.

Take appropriate precautions to prevent unauthorized access to these unencrypted private key files. If you need to roll back to a pre-4.1 version, see Restoring Unencrypted Private Keys.

Keep Reading

- Upgrading via UI
- Upgrading Manually
- Troubleshooting Cribl Edge Upgrades

7.1.2. Upgrading Edge Via UI

Transporting an Edge installation into the future.

This topic explains how to use the UI to upgrade, back up, or roll back a Cribl Edge single-instance or distributed deployment on Linux. If you prefer to do these operations manually – that is, at the command line – see the Upgrading Manually topic.

When you upgrade Cribl Edge via the UI, you must follow one of these upgrade paths:

Current Version	Upgrade Path
4.x	4.x
3.x	3.x through 4.x
2.x	2.x through 4.x
1.7.x or 2.0.x	2.x.x, then 3.x or 4.x
1.6.x or below	1.7.x, then 2.x.x, then 3.x or 4.x

Upgrading the Leader Node

To upgrade the Leader in a distributed deployment, go to **Settings**, then select **Upgrade** under **Global Settings**.

When upgrading, you can choose to either download upgrade packages from Cribl's content delivery network (CDN), or from a filesystem location that you specify in the form of a path.

The filesystem location must be accessible to the running Leader application, both in connectivity and permissions.

Selecting a Package Source

Here, you'll select between **CDN** and **Path** for your upgrade (in an on-prem, distributed deployment):

Package Source	Description	When You Select This Package Source
CDN	Downloads an installation package directly from	You'll see the currently installed version and the version available on the CDN. The Leader will always upgrade to the

Package Source	Description	When You Select This Package Source
	Cribl's content delivery network.	current CDN version when you click **Upgrade**.
Path	You specify the path to the installation package.	You must specify paths to installation packages that have compatible versions and architectures with your Leader and Edge Fleets. For example, a Worker Group installed on an ARM64 architecture will require an ARM64 installation package.

Use the buttons to choose the desired package source. Then, the UI will display settings and information appropriate for the package source you chose, as described in the respective **CDN** and **Path** sections below.

Configuring Settings for CDN Upgrade

If you select **CDN** as your upgrade method, you will see the following options:

- **Current CDN version**: Lists the latest version of Cribl Stream available on the CDN.
- **Leader**: This shows the currently installed version of Cribl Stream on the Leader. If a newer version is available, you will be able to use the **Upgrade to** button.
- **Stream Worker Groups**: Provides information and options for upgrading Stream Groups. See the Stream Documentation for more details.
- **Edge Fleets**: Displays **Upgrade Options**, including the deprecated **Enable Legacy Edge upgrades**.

Enable Legacy Edge Upgrades (Deprecated)

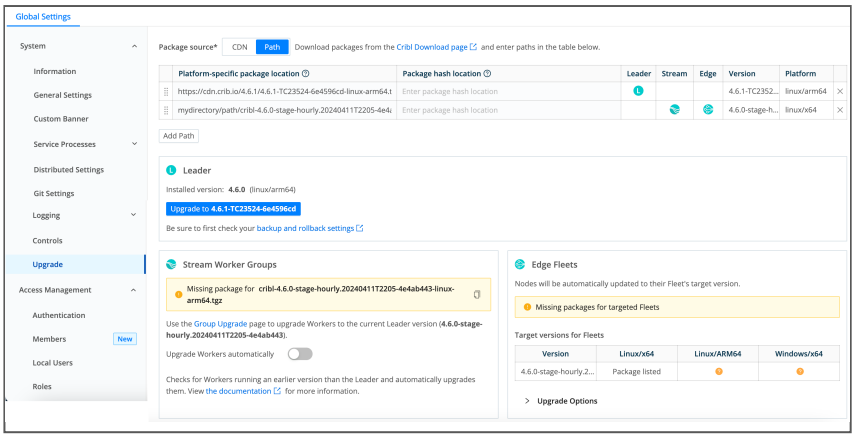
While this functionality still exists, it is deprecated and we do not recommend using it.

Defaults to `No`. Enabling Legacy upgrades removes the ability to automatically upgrade Nodes on a per-Fleet basis.

If you experience any upgrading issues with the 4.5.0+ Fleet upgrade functionality, consider toggling this to `Yes` to use the job-based upgrade framework and manually upgrade Fleets via the Fleet Upgrade UI and **Upgrade** button. You'll also need to toggle **Disable Jobs/Tasks** to `No`. See Disable Jobs/Tasks.

You may need to refresh the Global Settings page after enabling legacy upgrades to view the **Fleet Upgrade** tab.

Configuring Edge Settings for Path Upgrade



Upgrade Edge using Path

With **Path** as your upgrade method, you'll see the following settings and information in the **Package Source** table:

- **Platform-Specific Package Location**: Enter or paste the path to the Cribl installation package. This can be either of the following:
  - An HTTP URL, for example `https://cdn.cribl.io/dl/4.1.0/cribl-4.1.0-6979aea9-linux-x64.tgz`
  - A local filesystem path, for example `myfolder/directory/cribl-package.tgz`
- **Package Hash Location**: Enter either of the following:
  - An HTTP URL, for example `https://cdn.cribl.io/dl/4.1.0/cribl-4.1.0-6979aea9-linux-x64.tgz.sha256`
  - A local filesystem path to the hash that validates the package.

  Supports SHA-256 and MD5 formats. You can simply append `.sha256` to the contents of the **Platform-specific package location** field.
- **Leader**: Indicates the version the Leader will be upgraded to when an upgrade is available and you click the **Upgrade** button.
- **Stream**: The version in this row will be used to upgrade the Stream Worker Group, because it matches the Leader's version.
- **Edge**: The version in this row will be used to upgrade Edge Fleets with a matching target version.
- **Version**: Displays the version of the upgrade package you added in this row.
- **Platform**: Displays the platform architecture for the package in this row.

Select **X** to immediately delete a row – there is no confirmation prompt.

You can add multiple rows to this table to specify packages for different versions and platforms/architectures. To obtain the latest packages from https://cribl.io/download, use the drop-

> down list to specify each platform (for example, x64 versus ARM). When you stage these packages on your own servers, preserve the original file names.

Leader area: Shows you the currently installed Leader version and whether an upgrade is available. Check the backup and rollback settings before you upgrade.

Stream Worker Groups: Provides information and options for upgrading Stream Groups. See the Stream Documentation for more details.

Edge Fleets: Offers **Upgrade Options** for Edge. See Enable Legacy Edge Upgrades for more information.

Missing Package Links

When you see this warning, it means there are Edge Fleets that can't be upgraded due to a missing package path.

To add a missing installation package:

1. Click the ? icon, which copies the path to your clipboard.
2. Paste the copied package link into a new **Platform-specific package location** field.
3. Add the appropriate path directory to the beginning of the installation package path (where your Cribl upgrade packages are located).
   For example: `myfolder/directory/cribl-package.tgz`
4. Click **Save** to add the upgrade package path and resolve the warning.

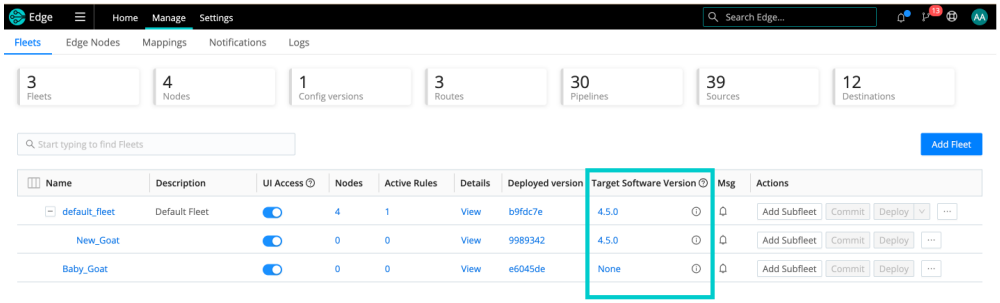
Upgrading Edge Nodes

You can upgrade the Nodes in a Fleet from the **Manage** > **Fleets** page. Upgrade options are on a per-Fleet basis:

- Choose **None** to never upgrade any of the Nodes in the Fleet.
- Select a target software version to upgrade all of the eligible Nodes in the Fleet to the selected version.

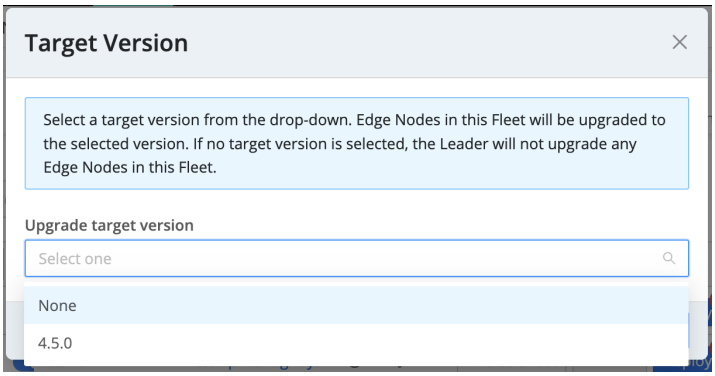
To upgrade the Nodes in a selected Fleet to a target version:

1. Go to **Manage** > **Fleets**.
2. In the **Target Software Version** column, click the current version (or **None**, if applicable).



New Fleet upgrade interface

3. Click the **Upgrade target version** drop-down and select a version.



Target version window

The next time the Node in the selected Fleet connects to the Leader, the Node will request the selected target version, and the Leader will upgrade the Node to that version.

Nodes in the selected Fleet will not upgrade to any future version until you select or change the target software version in the drop-down (even if you upgrade the Leader).

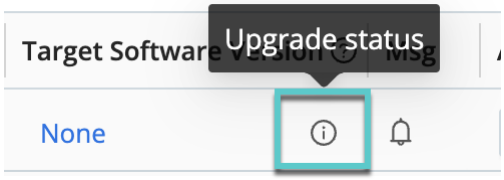
If you set the target software version to **None**, the Nodes in the selected Fleet will never upgrade. This lets you keep Edge Nodes at a specific version, even if you upgrade the Leader.

Subfleets

Subfleets do not inherit the **Target software version** from their parent Fleet, but you can set the target software version for each Subfleet using the same steps you used in the parent Fleet.

Fleet Upgrade Information

To view the upgrade-eligible Nodes in the selected Fleet and the Nodes that are on versions 4.4.4 and older, click the **Upgrade status** icon in the Target Software Version column.



Upgrade status icon

To find upgrade details for Nodes that were upgraded prior to 4.5.0, click the **Legacy Upgrade Jobs** tab.

Containerized Node Upgrades

If you're asking, "Why didn't my Kubernetes Node upgrade?" or perhaps, "Why *did* my Docker Node upgrade?", then you're in the right place.

Cribl Edge handles upgrades for containerized Nodes (4.5.0 and newer) differently than standalone or on-prem Nodes. Leader Nodes don't upgrade Edge Nodes that are running in a container (such as containerd, Docker, or Kubernetes), because Edge Nodes revert back to their Helm chart image after a Pod restarts. Instead, upgrade containerized Nodes manually. See Upgrading Kubernetes Edge Nodes.

Upgrades for containerized Nodes also work a little differently in 4.5.0 and newer compared to 4.4 and older. Consider these upgrade scenarios:

Scenario	Do the Nodes Upgrade?	Logic Explanation
A Fleet contains **Kubernetes / containerized** Nodes running version 4.5.0. The Fleet has a set **Target Software Version** of 4.5.2.	No	• The Nodes are running in a container. Containerized (Docker, Kubernetes, etc) Nodes will not be upgraded automatically. • Nodes on version 4.5.0 and later inform the Leader that they are containerized Nodes.
A Fleet contains **Kubernetes / containerized** Nodes running version 4.4. The Fleet has a set **Target Software Version** of 4.5.0.	Yes	• Nodes on version 4.4 and older do not have the ability to inform the Leader that they're containerized.

Greatgreen Goats: A Fleet Upgrading Use Case

To help you conceptualize automatic Fleet upgrades, this use case walks you through how the fictitious company Greatgreen Goats upgrades Nodes in their two Fleets.

Greatgreen Goats is a small but rapidly growing company focused on installing and maintaining solar energy systems at goat farms around the country.

They use Cribl Edge to manage data on their edge nodes, divided into two distinct Fleets:

- **Fleet 1: Linux Operations.** This Fleet contains 20 Nodes, all running Edge v.4.4.4. When the IT admin upgrades the Leader of Fleet 1 to 4.5, and they set the target upgrade version in the drop-down to **None**. Therefore, Nodes in Fleet 1 are never upgraded. They continue to remain at their current version, 4.4.4.
- **Fleet 2: Windows Monitoring.** This Fleet contains 15 Nodes. In this Fleet, the IT admin selects the current Leader version (in this case, 4.5) in the **Target upgrade version** drop-down. Now, when Nodes connect to the Leader, they pull down version 4.5 and upgrade to that version.

Fast forward four weeks: Cribl Edge has released a new version: 4.5.2. The Leader is now at version 4.5.2.

Since Fleet 1 does not have a target version selected, Nodes in Fleet 1 remain at their current version, 4.4.4. If the IT admin decides to change the **Target upgrade version** to match the Leader in Fleet 1, the Nodes in Fleet 1 will be upgraded to 4.5.2.

Nodes in Fleet 2 will remain on version 4.5, which was the Leader's previous version, until the admin selects the next **Target upgrade version** (4.5.2, the Leader's new version). Then, Nodes will request the new version the next time they connect to the Leader and automatically upgrade.

This approach gives the IT admin at Greatgreen the flexibility to upgrade Nodes in Fleets independently from the Leader. Greatgreen can now scale and upgrade their Fleets with more reliability, since Nodes are pulling the release down from the Leader (instead of the Leader creating an overwhelming number of upgrade jobs).

Legacy Upgrade Jobs

To view upgrades for Nodes older than 4.4.4, click the info icon in the **Target Software Version** column and click the **Legacy Upgrade Jobs** tab.

All Nodes in the selected Fleet that are on version 4.4.4 or older will be upgraded using the jobs framework. If legacy upgrade jobs exist, you can find them here in the Legacy Job Upgrades tab.

Backup and Rollback

When you initiate an upgrade through the UI, Edge first stores a backup of your current stable deployment. If the upgrade fails, then by default, Edge will automatically roll back to the stored backup package. You can

adjust this behavior at **Settings** > **Global Settings** > **System** > **General Settings** > **Upgrade & Share Settings**, using the following controls.

Edge can perform rollbacks only on Nodes that were running at least v.3.0.0 before the attempted upgrade.

Enable automatic rollback: Edge will automatically roll back an upgrade if the Edge server fails to start, or if the Edge Node fails to connect to the Leader. (Toggle to `No` to defeat this behavior.)

Rollback timeout (ms): Time to wait, after an upgrade, before checking each Node's health to determine whether to roll back. Defaults to `30000` milliseconds, i.e., 30 seconds.

Rollback condition retries: Number of times to retry the health check before performing a rollback. Defaults to `5` attempts.

Check interval (ms): Time to wait between health-check retries. Defaults to `1000` milliseconds, i.e., 1 second.

Backups directory: Specify where to store backups. Defaults to `$CRIBL_HOME/state/backups`.

Backup persistence: A relative time expression specifying how long to keep backups after each upgrade. Defaults to `24h`.

7.1.3. Upgrading Edge On The Command Line

Learn how to upgrade Edge Nodes manually using a command line interface (CLI).

In this guide, we'll cover benefits to upgrading manually and how to use the command line to upgrade your Edge Nodes.

A Word About Words

We will use "upgrading manually" and "on the command line / CLI" interchangeably in this article.

Why Upgrade Cribl Edge Manually?

With Edge, you have the option to upgrade manually or via the UI. Here are some use cases and benefits to using the command line to upgrade:

Flexibility and Control

Manually upgrading means you can pause and resume upgrades as needed, which gives you more granular control over the process.

You can also upgrade individual Nodes separately. This is useful when you're testing upgrades on a subset of Nodes before rolling out to the entire deployment.

Air-Gapped Environments and Custom Deployments

Consider upgrading via CLI when your organization can't connect to the internet for security or compliance reasons. You don't need to be online to upgrade on the command line.

When Cribl Edge is deployed in non-standard ways, such as via RPM package, within containers, or using a container orchestrator like Kubernetes, manual upgrades might be necessary.

Specific Upgrade Paths

When you upgrade Edge Nodes on a CLI, you don't have to worry about a version upgrade path - there are no restrictions on the versions you can upgrade to or from.

Upgrading a Standalone/Single-Instance

This path requires upgrading only the single/standalone Node:

1. Stop Edge.
2. Download the package on your instance of choice here.
3. Uncompress the new version on top of the old one, e.g., in the `/opt/` directory:

   ```
   tar xvzf cribl-<version>-<build>-<arch>.tgz
   ```

   - On some Linux systems, `tar` might complain with: `cribl/bin/cribl: Cannot open: File exists`. In this case, please remove `cribl/bin/cribl` and untar again.
   - If you have **custom functions** in `cribl/bin/cribl`, please move them to `$CRIBL_HOME/local/cribl/functions/` before untarring again.
4. Restart Edge.

Upgrading a Distributed Deployment

For a Distributed Deployment, the general order of upgrade is:

1. Upgrade the Leader Node.
2. Upgrade the Edge Nodes.
3. Commit and deploy the changes on the Leader.

For distributed environments with a second Leader configured for failover, this is the upgrade order:

1. Stop both Leaders.
2. Upgrade the primary Leader.
3. Upgrade the second Leader.
4. Upgrade each Edge Node, respectively.

Upgrading the Leader Node

1. Commit and deploy your desired last version. (This version will be your most recent checkpoint.)
   - Optionally, `git push` to your configured remote repo.
2. Stop Cribl Edge.
   - Back up the entire `$CRIBL_HOME` directory (recommended, but optional).

- Check that the Edge Nodes are still functioning as expected. In the absence of the Leader Node, they should continue to work with their last deployed configurations (optional).

3. Download the package on your instance of choice here.
4. Uncompress the new Edge version on top of the old one, e.g., in the `/opt/` directory:

```
tar xvzf cribl-<version>-<build>-<arch>.tgz
```

5. Restart Edge and log back in.
6. Wait for all the Edge Nodes to report to the Leader, and ensure that they are correctly reporting the last committed configuration version.

Cribl Edge's UI will not be available until the Edge version has been upgraded to match the version on the Leader. Errors will appear until the Edge Nodes are upgraded.

Upgrading the (Linux) Edge Nodes

These are the same basic steps as when upgrading a Single Instance, above:

1. Stop Cribl Edge on each Edge Node.
2. Download the package on your instance of choice here. If the Leader is on a prior release, see Cribl Past Releases for older packages.
3. Uncompress the new Edge version on top of the old one, e.g., in the `/opt/` directory:

```
tar xvzf cribl-<version>-<build>-<arch>.tgz
```

4. Restart Edge.

Upgrading Edge Nodes Installed via RPM

See Upgrading Edge via RPM.

Upgrading Kubernetes Edge Nodes

Edge Nodes deployed in Kubernetes can't be upgraded via the Leader. Instead, use our Helm charts to upgrade the Edge Nodes deployed in your Kubernetes cluster:

```
helm upgrade --install
```

Commit and Deploy Changes from the Leader Node

1. Ensure that newly upgraded Edge Nodes report to the Leader with their new software version.
2. Commit and deploy the newly updated configuration **only after all** Edge Nodes have upgraded.

Manual Rollback

Using CLI commands, it's possible to explicitly roll back an on-prem Leader, and Nodes, to an earlier release. This works much like an upgrade.

Explicit rollback might be necessary if an automatic rollback fails. Otherwise, Cribl recommends first considering other options – an upgrade, or working with Cribl Support – before performing a manual rollback.

Rollback can encounter these complications:

- Your current configuration might take advantage of dependencies not supported in an earlier release.

Do not manually roll back any Cribl Edge instance running in a container. Instead, locate, download, and launch a container image hosting the earlier version you want.

Rollback Outline

We assume that you are rolling back to a previously deployed version that you know to be stable in your environment. The broad steps are:

1. Ideally, link your deployment to a Git remote repo, and commit and push your Leader's configuration to that remote. The repo will provide a stable location from which to restore your config, if necessary.
2. Stop the Leader instance. (From `$CRIBL_HOME/bin/`, execute `./cribl stop`.)
3. Also create a local backup of your Leader's whole `$CRIBL_HOME` directory.
4. Obtain the installation package for the earlier release and platform you need. (See the Rollback Example.)
5. Uncompress the earlier version to your original deployed directory. You can do this from the command line or programmatically (see the Rollback Example).

If installing to your existing target directory fails, try the same mitigations we list above for upgrades.

6. In a distributed deployment, Nodes must not run a higher version than the Leader. Repeat steps 2, 3, and 5 above on all Nodes, substituting "Node" for "Leader." (On Nodes where you've integrated

Cribl Edge with systemd, stop the Cribl server with: `systemctl stop cribl-edge`.)

Rollback Example

While rollbacks can be partially scripted, discovering earlier installation packages cannot be automated – the initial steps here are manual:

1. Stop and back up the Leader. (Follow steps 1–3 in the Rollback Outline.)
2. Open the **Cribl Past Releases** page: https://cribl.io/download/past-releases/.
3. Locate the earlier version that you want to restore.
4. Use the adjacent drop-down to select your target platform.
5. Right-click (or `Ctrl`-click) the corresponding **Download** button, and copy its URL to your clipboard. In this example, we've copied:
   `https://cdn.cribl.io/dl/4.1.0/cribl-4.1.0-6979aea9-linux-arm64.tgz`
6. Swap that URL into an untar command like the following. Run this command from the parent (typically `/opt/`) of your existing `$CRIBL_HOME` directory:
   `[sudo] curl -Lso - https://cdn.cribl.io/dl/4.1.0/cribl-4.1.0-6979aea9-linux-arm64.tgz | tar zxv`
7. Repeat the preceding steps to adjust all Nodes to a compatible version.

Cribl University offers a course titled Edge Administration that provides an illustrated overview of administering Cribl Edge. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

7.1.4. Troubleshooting Edge Upgrades

In case of upgrading emergency, break glass.

While we expect you'll have an easy time upgrading Cribl Edge, here is the troubleshooting information we've curated in case you do experience a hiccup:

Direct Upgrades

Cribl Edge does **not** support direct upgrades from any pre-GA version (such as a Cribl-provided test candidate) to a GA version. To get the GA version running, you must perform a new install.

Upgrading to v.3.5.4 or Later

- If you're upgrading to v.3.5.4 or later, all Edge Nodes will need to be on the same version as the Leader. Leaders running v.3.5.4 and later test whether Edge Nodes are running a compatible version before deploying configs that could break Nodes' data flow. The Leader will prompt you to upgrade these Nodes as needed. For details, see the Edge 3.5.4 release notes.
- Version 3.5.4 is also a compatibility breakpoint for the Cribl HTTP Source and Destination, and for the Cribl TCP Source and Destination. When running on Cribl Edge 3.5.4 and later, these two Sources can send data only to Workers running v.3.5.4 and later, and these two Destinations can receive data only from Nodes running v.3.5.4 and later. When running on Edge 3.5.3 and earlier, these four integrations can similarly interoperate only with Nodes running v.3.5.3 and earlier.

7.2. Notifications Overview

Notifications alert Cribl Edge admins about issues that require their immediate attention.

This page describes the uses for Cribl Edge Notifications, their license requirements, and how to configure them.

Types of Notifications

In Cribl Stream (LogStream) 3.1 or later, and all Cribl Edge versions, you can configure Notifications about:

- Sources and Collectors that report abnormally high or low data flow rates. For details, see Source-State Notifications.
- Sources and Collectors that report no data flow. For details, see Source-State Notifications.
- Destinations experiencing backpressure. For details, see Destination-State Notifications.
- Destinations approaching their persistent queue threshold. For details, see Destination-State Notifications.
- Destinations that report errors. For details, see Destination-State Notifications.
- Pending expiration of a Cribl Edge license. For details, see License-Expiration Notifications in Cribl.

Notifications are also sent as events to Cribl Edge's internal logs – both application-wide, and with a filtered view available on affected Sources and Destinations. Cribl Edge stores these application-wide logs in `notifications.log` on the Leader Node. The Leader Node is also responsible for sending all Notifications.

License Requirements

Notifications require an Enterprise or Standard license. Without an appropriate license, the configuration options described below will be hidden or disabled in Cribl Edge's UI.

Notifications and Targets

A Notification target specifies the delivery method for a Notification. Every Notification requires one or more targets.

Available target types include:

- Webhook
- PagerDuty integration

- Slack
- AWS SNS
- Email

For details, see Notification Targets.

By default, any Notification that you configure will have a target of `System Messages`. When a Notification condition is triggered, Cribl Edge will add an indicator on the top nav's **Messages** button. Click this button to view details in the **Messages** drawer.

Notifications and RBAC

Notifications work with Cribl Edge's role-based access control. For users with non-administrative permissions, their assigned Roles and Policies determine the Worker Groups on which they can view Notification messages, configure Notifications, and configure targets.

Implementing Notifications

Implementing a Notification is a three-step process:

- Create a Notification about a Source state, Destination state, or pending license expiration.
- Add or create a Notification target.
- Test the target to be sure it will work when needed.

You can create either the Notification or the target first — the order doesn't matter.

To create a Notification about a Source or Destination state, click **Add Notification** in the Source or Destination modal, **Notifications** tab.

To create a license-expiration Notification, click **Add Expiration Notification** in the **Global Settings** > **Licensing** UI.

General

ID: Provide a unique ID for the Notification in this section. Notifications are enabled by default, but you can disable the Notification by setting **Enabled** to `No`.

Configuration

Define the triggering condition for the Notification in this section. Conditions vary by Notification type, so consult one of the following topics for the relevant details:

- Source-State Notifications
- Destination-State Notifications
- License-Expiration Notifications

Click **Add Target** to add a preexisting target. Click **Create Target** to create a new target.

If you select or add an email Notification target, an additional set of configuration options appear. See Configuring Email Notifications for more information.

Metadata

Metadata fields are user-defined fields included in the notification payload. All Notification types can contain metadata.

Click **Add field** here to add custom metadata fields to your Notifications in the form of key-value pairs:

- **Name**: Enter a name for this custom field.
- **Value**: Enter a JavaScript expression that defines this field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

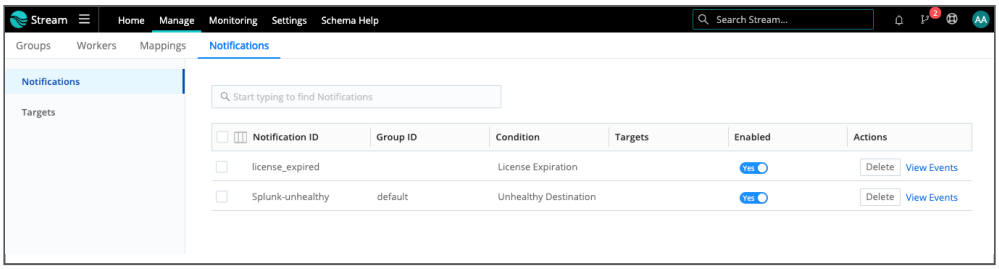
Once you've saved your Notifications, you can see Notification events specific to this Destination on the Destination config modal's **Events** tab. (When you set Source-state Notifications, a corresponding **Events** tab is available on Sources' and Collectors' config modals.) For a comprehensive view of all Notification events, see the systemwide Events Tab.

Managing Notifications

You can manage existing Notifications and targets from **Manage** > **Notifications**. You can also reach this UI by clicking an existing Notification in a Source or Destination modal or in the **Global Settings** > **Licensing** UI.

Notifications Tab

This tab lists all your configured Source-state and Unhealthy Destination Notifications, across all integrations, along with any configured license-expiration Notifications. You can't create new Notifications here, but you can disable or delete existing Notifications. For any individual Notification, click **View Events** to see events that have triggered the Notification.

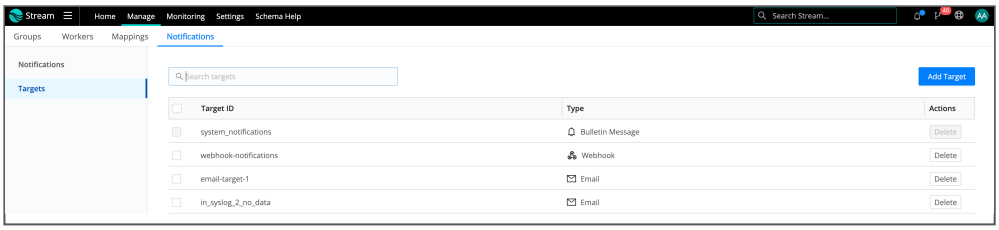


Notifications tab

To modify a Notification, click anywhere on its row.

Targets Tab

This tab is where you centrally configure and manage targets that are available across Cribl Edge – for all Sources, Destinations, and license-based Notifications. See Notification Targets for details.



Targets tab

To create a new target, click **Add Target**. To delete a target, click **Delete** in the appropriate row.

7.2.1. Source-State Notifications

In Cribl Edge 3.5 and above, you can configure Notifications on Sources and Collectors to trigger under these conditions:

- High Data Volume
- Low Data Volume
- No Data

Read on for details about these conditions and how to configure appropriate Notifications.

High Data Volume

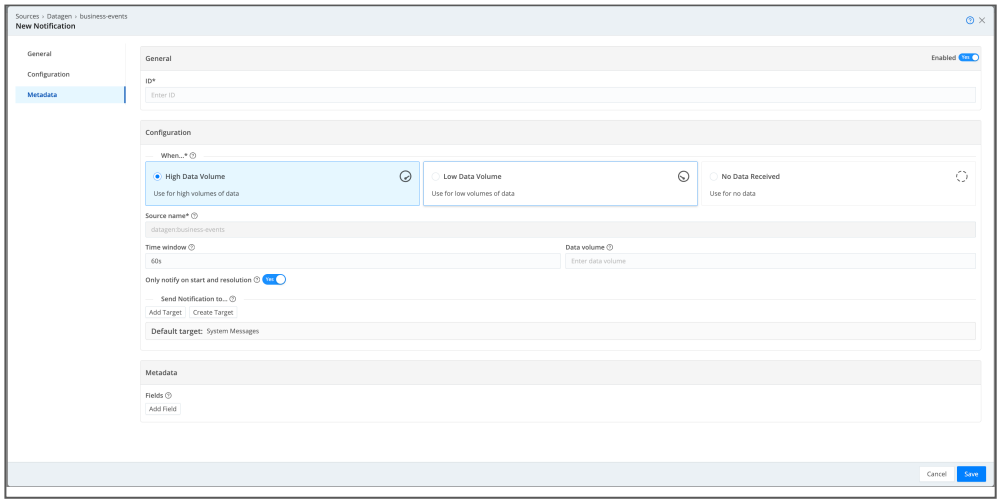
Cribl Edge will generate a Notification when incoming data over your configured **Time window** exceeds your configured **Data volume** threshold. This selection exposes the following fields:

Notification targets: The **Default target** is always locked to `System Messages`.

Source name: This field is locked to the Source on which you're setting this Notification.

Time window: This field's value sets the threshold period before the Notification will trigger. The default `60s` will generate a Notification when the Source has reported the trigger condition over the past 60 seconds. To enter alternative numeric values, append units of `s` for seconds, `m` for minutes, `h` for hours, and so forth.

Data volume: Enter the threshold above which a Notification will trigger. Accepts numerals with units like `KB`, `MB`, and so forth. For example: `4GB`.



Configuring a high data volume Notification

Low Data Volume

Select the `Low Data Volume` tile to trigger Notifications when incoming data over your configured **Time window** is lower than your configured **Data volume** threshold.

This selection exposes the same additional fields as High Data Volume, except that here, the **Data volume** value defines a **floor** below which the Notification will trigger.

No Data

Select the `No Data Received` tile to trigger Notifications when the Source or Collector ingests zero data over your configured **Time window**.

This selection exposes the same additional fields as High Data Volume, except it omits the **Data volume** field. With no data entering the Source, there is no threshold to configure.

Source-State Notifications for Webhook Targets

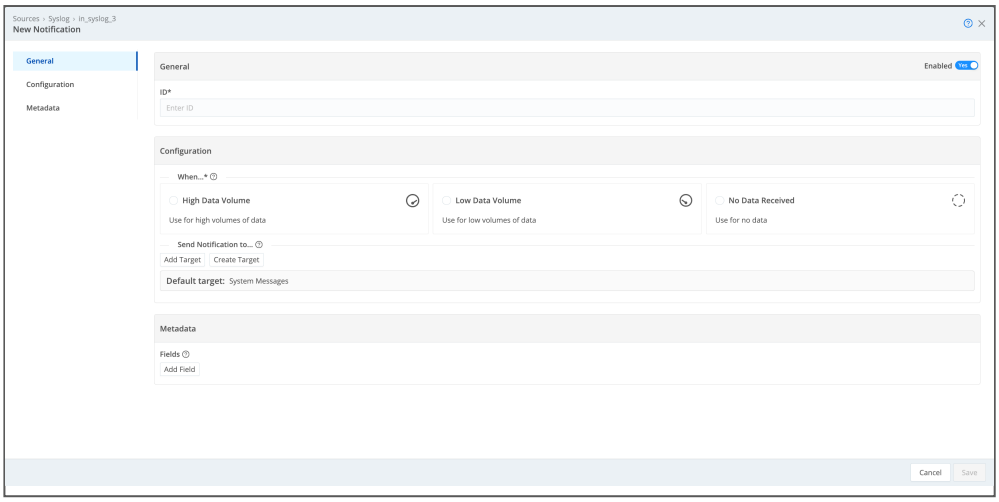
If you are sending a Source-state Notification to a Webhook Notification target, you can include a variety of expression fields in the target's **Source expression**. For more information, see:

- Fields Common to All Notification Types
- Source High Data Volume
- Source Low Data Volume
- Source No Data Received

Configuring Source Notifications

To configure a Source-state Notification:

1. Configure and save the Source.
2. Access this Source's **Notifications** tab by one of the following methods:
    - Click the **Notifications** button on the **Manage...Destinations** page's appropriate row.
    - Reopen the Source's config modal and click its **Notifications** tab.
3. Click **Add Notification** to access the **New Notification** modal shown below.



Configuring a Source Notification

General

ID: Enter a unique ID for this Notification. Notifications are enabled by default, but you can disable the Notification by setting **Enabled** to `No`.

Configuration

When: Select one of the following Notification tiles:

- **High Data Volume**
- **Low Data Volume**
- **No Data Received**

You can create multiple Notifications for the same Source, but you must configure them separately.

Send notification to: Click **Add Target** to send this Notification to additional targets. You can add multiple targets.

- Use the resulting **Notification targets** drop-down to select any target you've already configured.
- Click **Create Target** to configure a new target.

See Notification Targets for details.

Default target is always locked to `System Messages`.

Source name: This field is locked to the Source on which you're setting this Notification.

Time window: This field's value sets the threshold period before the Notification will trigger. The default `60s` will generate a Notification when a Destination or Source has reported the trigger condition over the past

60 seconds. To enter alternative numeric values, append units of `s` for seconds, `m` for minutes, `h` for hours, and so forth.

Only notify on start and resolution: When this option is set to `Yes`, Cribl Edge will send a Notification at the onset of the triggering condition. On resolution, it will send a second Notification.

If you don't enable this option and a Source-state Notification's trigger condition persists beyond your configured **Time window**, Cribl Edge will send a new Notification, once per **Time window** interval.

Metadata

You can enter user-defined fields called *metadata*, which Cribl Edge includes in the notification payload. See Metadata for more information.

7.2.2. Destination-State Notifications

In Cribl Edge 3.5 and above, you can configure Notifications on Destinations that will trigger under these conditions:

- Destination Backpressure Activated
- Persistent Queue Usage
- Unhealthy Destination

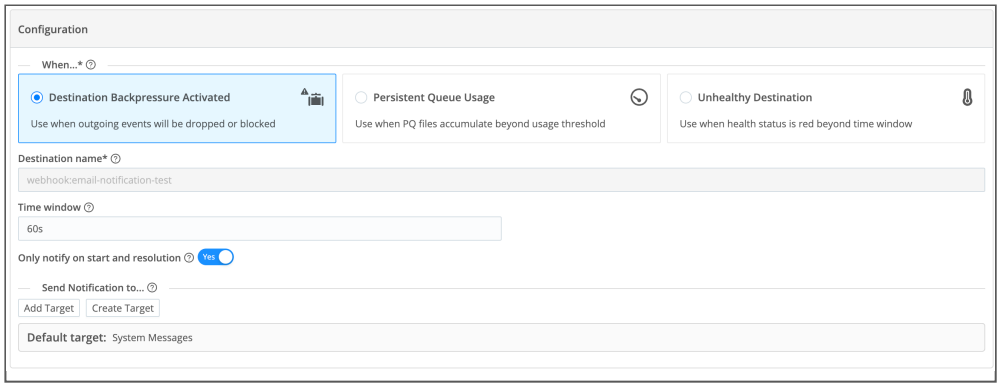
Read on for details about these conditions and how to configure appropriate Notifications.

Destination Backpressure Activated

Cribl Edge will generate a Notification when one of the following events occurs:

- The Destination's **Backpressure behavior** is set to `Block` or `Drop`, and backpressure causes outgoing events to block or drop.
- The Destination's **Backpressure behavior** is set to Persistent Queue, but its **Queue-full behavior** is set to either `Block` or `Drop new data`, and a filled queue causes the Destination to to block or drop outgoing events.

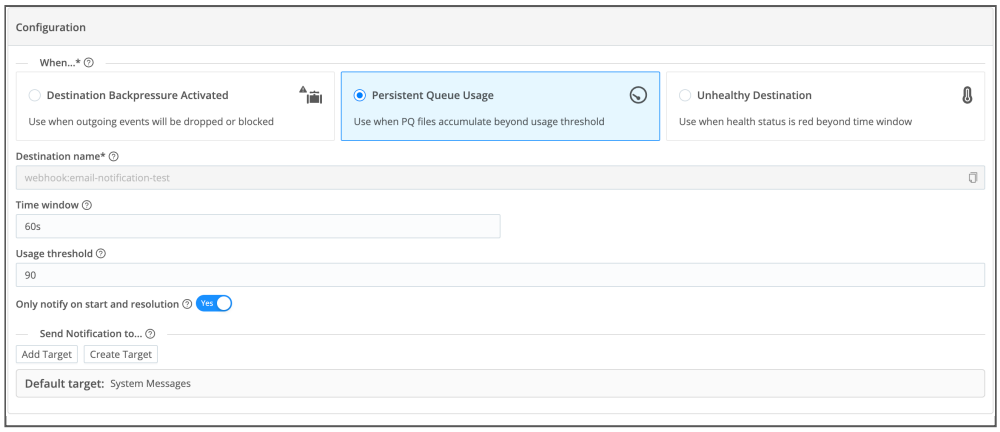
The threshold for the Notification to trigger is: Cribl Edge detected a blocked or dropped state during ≥ 5% of the trailing **Time window** that you configure in the Configuring Destination Notifications.



Backpressure Notification

Persistent Queue Usage

Cribl Edge will generate a `Persistent Queue usage has surpassed <threshold>%` Notification when the PQ accumulates files past the `<threshold>` percentage of capacity that you set in the **Usage threshold** field. This field appears only when you configure a Notification for **Persistent Queue Usage**.



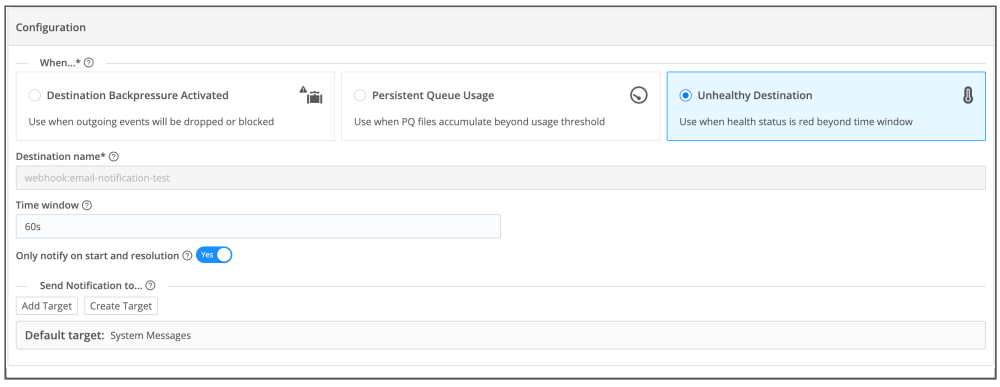
Persistent Queue Notification

Unhealthy Destination

Cribl Edge will generate a `Destination <name> is unhealthy` Notification when the Destination's health has been in red status (as indicated on the UI's Monitoring page) over the trailing **Time window** that you configure in Configuring Destination Notifications.

The algorithm has slight variations among Destination types, but red status generally means that ≥ 5% of health checks, aggregated over the **Time window**, reported one of the following conditions:

- An error inhibiting the Destination's normal operation, such as a connection error.
- For multiple-output Destinations like Splunk Load Balanced or Output Router, > 50% of the Destination senders are in an error state.



Destination Unhealthy Notification

Destination-State Notifications for Webhook Targets

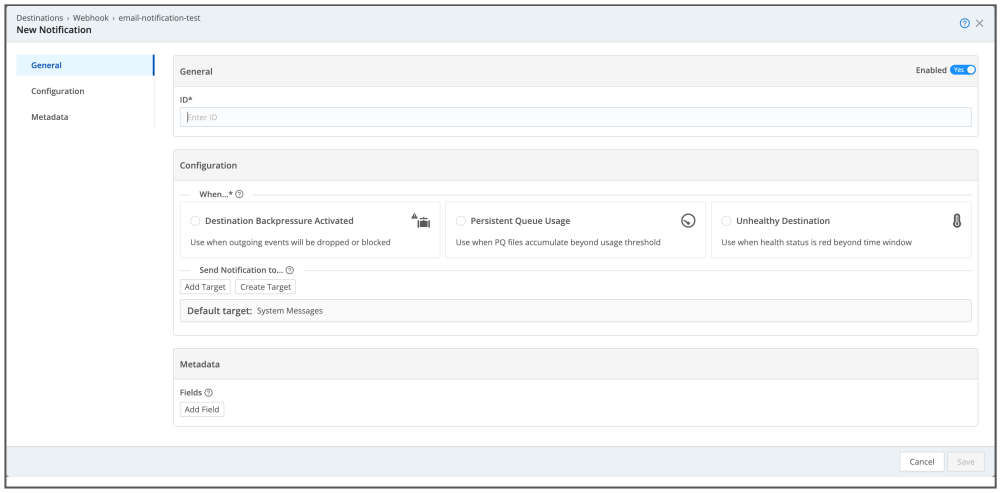
If you are sending a Destination-state Notification to a Webhook Notification target, you can include a variety of expression fields in the target's **Source expression**. For more information, see:

- Fields Common to All Notification Types
- Unhealthy Destination
- Destination Backpressure Activated

Configuring Destination Notifications

To configure a Destination-state Notification:

1. Configure and save the Destination.
2. Access this Destination's **Notifications** tab by one of the following methods:
    - Click the **Notifications** button on the **Manage...Destinations** page's appropriate row, or
    - Reopen the Destination's config modal and click its **Notifications** tab.
3. Click **Add Notification** to access the **New Notification** modal shown below.



Configuring a Destination Notification

General

ID: Enter a unique ID for this Notification. Notifications are enabled by default, but you can disable the Notification by setting **Enabled** to `No`.

Configuration

When: Select one of the following Notification tiles:

- **Destination Backpressure Activated**
- **Persistent Queue Usage**
- **Unhealthy Destination**

You can set up multiple Notifications for the same Destination, but you must configure them separately.

Send Notification to: Click **Add Target** to send this Notification to additional targets. You can add multiple targets.

- Use the resulting **Notification targets** drop-down to select any target you've already configured.
- Click **Create Target** to configure a new target.

See Notification Targets for details.

Default target is always locked to `System Messages`.

Destination name: This field is locked to the Destination on which you're setting this Notification.

Time window: This field's value sets the threshold period before the Notification will trigger. The default `60s` will generate a Notification when a Destination or Source has reported the trigger condition over the past 60 seconds. To enter alternative numeric values, append units of `s` for seconds, `m` for minutes, `h` for hours, and so forth.

Only notify on start and resolution: When this option is set to `Yes`, Cribl Edge will send a Notification at the onset of the triggering condition and a second Notification to report its resolution.

If you don't enable this option and a Destination-state Notification's trigger condition persists beyond your configured Time window, Cribl Edge will send a new Notification, once per **Time window** interval.

Metadata

You can enter user-defined fields called *metadata*, which Cribl Edge includes in the notification payload. See Metadata for more information.

7.2.3. License-Expiration Notifications

To prevent interruptions in data throughput, you can configure a Notification that will be triggered two weeks before your Cribl Edge paid license expires, and then again upon expiration. (If the two-week Notification is cleared from the **Messages** tab between those dates, but the license has not been extended, it will trigger again.)

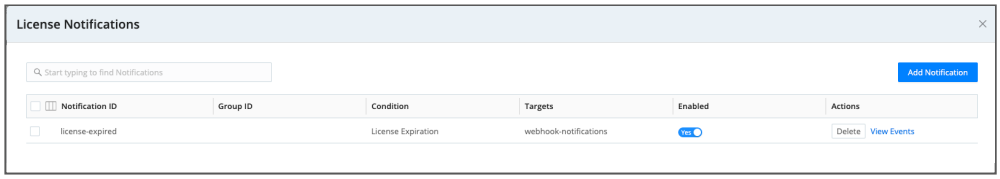
License-Expiration Notifications for Webhook Targets

If you are sending a license-expiration Notification to a Webhook Notification target, you can include a variety of expression fields in the target's **Source expression**. For more information, see:

- Fields Common to All Notification Types
- License Expiration

Configuring License-Expiration Notifications

1. From the top nav, select **Settings** > (**Global Settings** >) **Licensing**.
2. Click **Add Expiration Notification** to access the **License Notifications** modal.



License Notifications modal

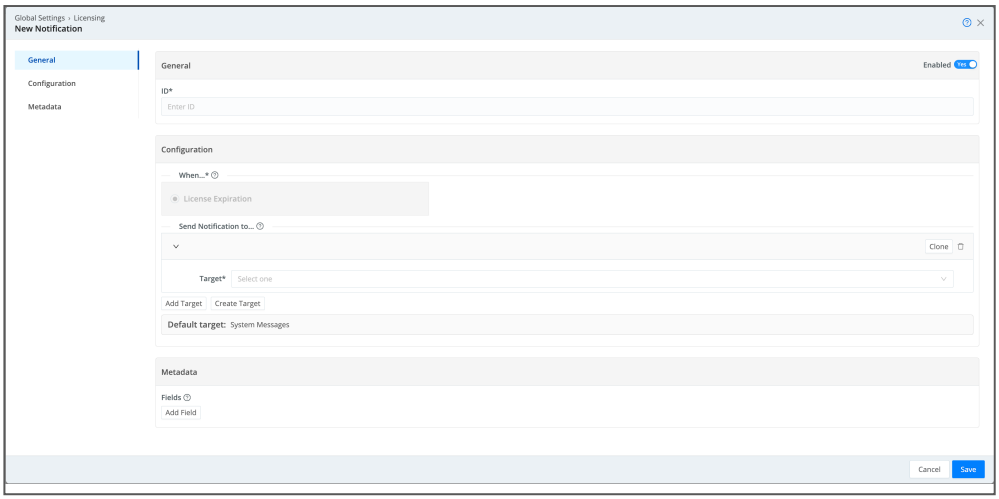
This modal shows existing license Notifications.

To modify a license-expiration Notification, click anywhere on its row.

To delete a Notification, click **Delete**.

To view license-expiration events, click **View Events**.

Click **Add Notification** to to access the **New Notification** modal shown below.



Configuring a license expiration Notification

This **New Notification** modal provides General, Configuration, and Metadata tabs.

General

ID: Enter a unique ID for this Notification. Notifications are enabled by default, but you can disable the Notification by setting **Enabled** to `No`.

Configuration

When: This modal's triggering condition is locked to `License Expiration`.

Send notification to: This section contains a list of the targets receiving this Notification. The **Default target** is always locked to `System Messages`.

Click **Add Target** for each additional existing target that you want to send this Notification to. Click **Create Target** to create a new target for the Notification.

Metadata

You can enter user-defined fields called *metadata*, which Cribl Edge includes in the notification payload. See Metadata for more information.

7.2.4. Email Notifications

If you're a Cribl Edge admin, email notifications make it easy to receive alerts about any operational issues that require your attention, such as a particular Source or Destination condition or a pending license expiration.

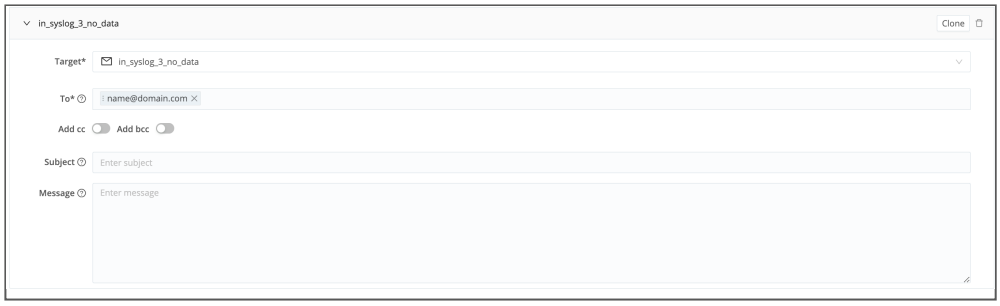
An email Notification requires two things — a configured Notification and an email Notification target. Cribl.Cloud organizations using an Enterprise license also have access to a preconfigured email Notification target.

Email Notifications do not have an unsubscribe option. Recipients who do not want to receive particular email Notifications should contact their Cribl Edge admins.

Email Notifications are sent on a no-reply basis.

Configuring Email Notifications

When you create a Notification for an email target, specify the recipients of the message, the subject line, and the contents of the message.



Email Notification with target selected

Target: The ID of the email target to which you want to send the Notification.

Click **Add target** to add an existing target.

Click **Create target** to create a new target. For information on configuring email Notification targets, see Email Notification Targets.

If the Notification already has a designated target, you can change the selection by clicking the drop-down.

When an email notification target is selected, the following additional fields appear:

To: The email address of the recipient.

Add cc: When enabled, reveals a field where you can enter the addresses of additional recipients.

Add bcc: When enabled, reveals a field where you can enter the addresses of additional recipients that do not appear in the Notification email.

Cribl Edge does not limit the total number of recipients for a Notification, but your email service might set a limit.

Subject: The subject line of the email Notification. You can use variables in the subject line.

Message: The content of an email Notification. You can use variables in the body of the email message.

Email Notification Variables

Email variables are placeholders in the email template that get replaced with actual values when the email is sent. These variables can be:

- General-use variables, like `condition`, `worker_group`, or `timestamp`
- Special-purpose variables

You can use a variety of general-use and special-purpose variables in the subject or message body of an email Notification. Insert a variable name between two braces preceded by a `$`. For example: `${cribl_notification}`.

General-Use Email Variables

Variable	Description
`workspace`	Workspace name (Cribl.Cloud only).
`organization`	Organization ID (Cribl.Cloud only).
`timestamp`	Timestamp when the email is sent. For example: `2019-08-04 18:22:24 UTC`.
`cribl_notification`	User-defined notification ID.

Special-Purpose Email Variables

Variable	Usage	Example
`input`	Type and name of a Source or Collector.	`syslog:in_syslog_1`
`output`	Type and name of a Destination.	`webhook:out_webhook_1`
`name`		Unique ID of a Source or Destination.
`bytes`	Quantity of data triggering the specified Notification.	`0`
`starttime`	Start time of an Event (in Epoch seconds).	`1706747185`
`endtime`	End time of an Event (in Epoch seconds).	`1706747294`
`health`	Health metric of the specified Source or Destination.	`0`
`_raw`	The `_raw` field of the Event triggering the Notification.	`_raw (Source ${name} in group ${__workerGroup} traffic volume greater than ${dataVolume} in ${timeWindow})`
`backpressure_type`	Backpressure behavior.	`1=BLOCKING, 2=DROPPING`
`queue_usage`	Percentage of capacity set in the **Usage threshold** field for a **Persistent Queue Usage** Notification.	`90`

Troubleshooting Email Notifications

The section details the troubleshooting steps you can take if an email notification fails to reach its intended recipient.

Test the Email Notification Target

An email notification can fail if the target is misconfigured. You can test your email notification target by following this procedure:

1. Open the target (in **Manage** > **Notifications** > **Targets**).

2. Click **Test Target**.
3. Add one or more email addresses to the **Test Target** modal and click **Send Test Email**.
4. Check the designated inbox to verify receipt of the test message. If it does not arrive in the designated inbox, review the target configuration.

Check Notification Service Logs

A failed email notification leaves a log entry. You can examine logs by following this procedure:

1. Select **Monitoring** > **Logs**.
2. Open the **Logs** drop-down and select **Leader** > **Notifications Service**.
3. Examine any logs that have errors.

Check Notifications

Cribl Edge stores Notifications. You can check them by following this procedure:

1. Select **Monitoring** > **Notifications**.
2. Select the **cribl_notification** field.
3. Search for **cribl_notification** fields corresponding to the Notification ID of the failed Notification.

7.2.5. Notification Targets

A Notification target specifies the delivery method for a Notification. Every Notification requires one or more targets.

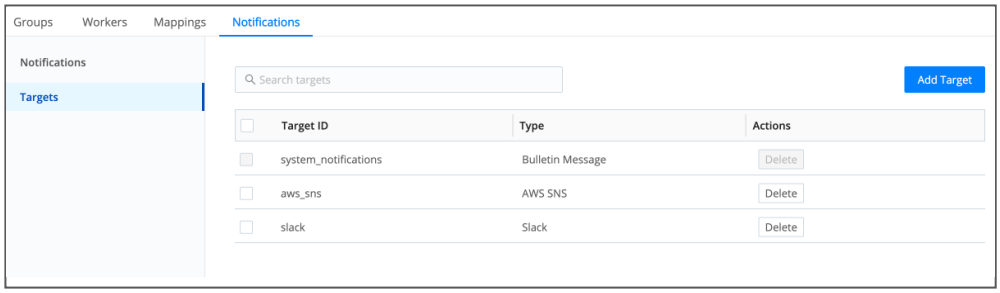
Available target types include:

- Webhook
- PagerDuty integration
- Slack
- AWS SNS
- Email

Adding a Notification Target

To add a new Notification target from the Manage Notifications page's **Targets** tab:

1. Click **Add Target** to open the **New Target** modal shown below.
2. Give this target a unique **Target ID**.
3. Select the desired **Target type**.



Add targets on the **Targets** tab

Then configure the target as described in the appropriate page:

- Webhook
- PagerDuty
- Slack
- AWS SNS
- Email

When you create a Notification target in Cribl Stream, Edge, or Search, it will also be available to you in the other products. For example, if you create a Notification target in Cribl Stream, you can also access it in Cribl Edge and Cribl Search.

Notifications require an Enterprise or Standard license. Without an appropriate license, target configuration options will be hidden or disabled in Cribl Edge's UI.

Managing Notification Targets

You can manage targets by selecting **Manage** > **Notifications** > **Targets**.

Click any existing Notification target to open a modal where you can manage the target, using buttons at the bottom of the UI.

Manage notification targets

Click **Delete Target** to delete an existing Notification target. You can also select multiple targets for deletion, using the check boxes on the left side of the modal.

Click **Clone Target** to copy an existing target. A modal will open so you can modify the configuration of the target you started with.

To edit any target's definition in a JSON text editor, click **Manage as JSON** at the bottom of the **Notifications** > **Targets** modal. You can directly edit multiple values, and you can use the **Import** and **Export** buttons to copy and modify existing target configurations as `.json` files.

Some targets have additional controls for target configuration. See the documentation for the target of interest for more information.

7.2.5.1. Webhook Notification Targets

With this option, you can send Cribl Edge Notifications to an arbitrary webhook. Select **Webhook** in the **New Target** modal to expose multiple left tabs, with the following configuration options:

General Settings

Target ID: Enter a unique ID used to identify the target. This will show in the **Target ID** column of the **Targets** tab. You can't change it later, so make sure you like it.

Configuration

The added options that appear on this first left tab are:

URL: The endpoint that should receive Cribl Edge Notification events.

To proxy outbound HTTP/S requests, see System Proxy Configuration.

Method: Select the appropriate HTTP verb for requests: `POST` (the default), `PUT`, or `PATCH`.

Format: Specifies how to format Notification events before sending them to the endpoint. Select one of the following:

- `NDJSON` (newline-delimited JSON, the default).
- `JSON Array`.
- `Custom`, which exposes these additional fields:
    - **Source expression**: JavaScript expression whose evaluation shapes the event that Cribl Edge sends to the endpoint – for example: `notification=${_raw}`. For other fields you can use, see Expression Fields. If empty, Cribl Edge will send the full notification event as stringified JSON.
    - **Drop when null**: Toggle to `Yes` if you want to drop events where the above **Source expression** evaluates to `null`.
    - **Event delimiter**: Delimiter string to insert between individual events. Defaults to newline character (`\n`).
    - **Content type**: Defaults to `application/x-ndjson`. You can substitute a different content type for requests sent to the endpoint. This entry will be overridden by any content types set in this modal's **Advanced Settings** tab > **Extra HTTP Headers** section.

- **Batch expression**: Expression specifying how to format the payload for each batch. Defines a wrapper object in which to include the formatted events, such as a JSON document. This enables requests to APIs that require such objects. To reference the events to send, use the `${events}` variable. An example expression to send the batch inside a JSON object would be: `{"items" : [${events}] }`.

Expression Fields

When building the Source expression, you can use the following fields:

Fields Common to All Notification Types

- `starttime`: Beginning of the time bucket where this condition was reported. All Notifications have this field.
- `endtime`: End of the time bucket where this condition was reported. All Notifications have this field.
- `_time`: Timestamp when this Notification was created. All Notifications have this.
- `cribl_host`: Hostname of the (physical or virtual) machine on which this Notification was created. All Destination and Source Notifications have this.
- `cribl_notification`: Configured name/ID of this Notification. All Destination and Source Notifications have this.
- `origin_metadata`: Object containing metadata about the Notification's origin, with the following fields for all Destination Notifications:
  - `type`: "output".
  - `id`: ID of the affected Destination.
  - `subType`: Destination's type (where applicable).
- `origin_metadata`: Object containing metadata about the Notification's origin, with the following fields for all Source Notifications:
  - `type`: "input".
  - `id`: ID of the affected Source.
  - `subType`: Source's type (where applicable).

Unhealthy Destination

- `health`: Numeric value where `0`=green, `1`=yellow, `2`=red.
- `output`: Output ID of the affected Destination.
- `_raw`: "Destination `${output}` [in group `${__worker_group}`] is unhealthy".
- `_metric`: "health.outputs".

Destination Persistent Queue Usage

- `output`: Output ID of the affected Destination.
- `timeWindow`: The interval at which notifications will be repeated.
- `usageThreshold`: The minimum persistent queue utilization (stated as a percentage) that will trigger Notifications.
- `_metric`: "system.pq_used".
- `usage`: The persistent queue utilization (stated as a percentage) that triggered the Notification.

Destination Backpressure Activated

- `backpressure_type`: `1` for Block, `2` for Drop.
- `output`: Output ID of the affected Destination.
- `_raw`: "Backpressure ([dropping|blocking]) is engaged for destination `${output}` [in group `${__worker_group}`]".
- `_metric`: "backpressure.outputs".

Source High Data Volume

- `health`: Numeric value where `0`=green, `1`=yellow, `2`=red.
- `bytes`: Number of bytes received in the time bucket.
- `input`: Input ID of the affected Source.
- `_raw`: "Source `${input}` [in group `${__worker_group}`] traffic volume greater than `${dataVolume}` in `${timeWindow}`".
- `_metric`: "total.in_bytes".

Source Low Data Volume

- `health`: Numeric value where `0`=green, `1`=yellow, `2`=red.
- `bytes`: Number of bytes received in the time bucket.
- `input`: Input ID of the affected Source.
- `_raw`: "Source `${input}` [in group `${__worker_group}`] traffic volume less than `${dataVolume}` in `${timeWindow}`".
- `_metric`: "total.in_bytes".

Source No Data Received

- `health`: Numeric value where `0`=green, `1`=yellow, `2`=red.
- `_time`: Timestamp when this Notification was created.
- `input`: Input ID of the affected Source.
- `_raw`: "Source `${input}` [in group `${__worker_group}`] had no data for `${timeWindow}`".
- `_metric`: "total.in_bytes"

License Expiration

- `severity`: One of "warn" or "fatal".
- `title`: One of: "License expiring soon, data will stop flowing." Or: "License has expired. Data flow has been stopped.".
- `text`: One of: "License will expire on `${expirationDate}`, no external inputs will be read" after that time. Please contact sales@cribl.io to renew your license." Or: "License has expired."

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.
- **Basic**: In the resulting **Username** and **Password** fields, enter HTTP Basic authentication credentials.

Post-Processing Settings

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).

- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates not authorized by a CA in the **CA certificate path**, nor by another trusted CA (for example, the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Keep alive: By default, Cribl Edge sends `Keep-Alive` headers to the remote server and preserves the connection from the client side up to a maximum of 120 seconds. Toggle this off if you want Cribl Edge to close the connection immediately after sending a request.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB, but you can set this limit to as high as 500 MB (512000 KB).

High values can cause high memory usage per Edge Node, especially if a dynamically constructed URL causes this target to send events to more than one URL. The actual request body size might exceed the specified value because the target adds bytes when it writes to the downstream receiver. Therefore, we recommend that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass to all events as additional HTTP headers. Values will be sent encrypted. You can also add headers dynamically on a per-event basis in the `__headers` field; headers added by this method take precedence over headers defined in the **Extra HTTP headers** table.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Used to declare headers that are safe to log as plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

7.2.5.2. PagerDuty Notification Targets

This option sends Cribl Edge Notifications to PagerDuty, a real-time incident response platform, using Cribl Edge's native integration with the PagerDuty API. Select Select **PagerDuty** in the **New Target** modal to expose the following additional options on the modal's (single) **General Settings** left tab:

General Settings

Target ID: Enter a unique ID used to identify the target. This will show in the **Target ID** column of the **Targets** tab. It can't be changed later, so make sure you like it.

Configuration

Routing key: Enter your 32-character Integration key on a PagerDuty service or global ruleset.

Group: Optionally, specify a PagerDuty default group to assign to Cribl Edge Notifications.

Class: Optionally, specify a PagerDuty default class to assign to Cribl Edge Notifications.

Component: Optionally, a PagerDuty default component value to assign to Cribl Edge Notification. (This field is prefilled with `logstream`.)

Severity: Set the default message severity for events sent to PagerDuty. Defaults to `info`; you can instead select `error`, `warning`, or `critical`. (Will be overridden by the `__severity` value, if set.)

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

7.2.5.3. Slack Notification Targets

You can send notifications to a Slack channel, using Slack's Incoming Webhooks.

You can also use a Webhook Destination to send specific event data to Slack for a team's attention. For more information, see our Slack/Webhook integration topic.

First, in your Slack workspace, use a Slack app to enable incoming webhooks.

Then, in the **New Target** modal, click **Slack** to expose the following additional options on the modal's (single) **General Settings** left tab:

General Settings

Target ID: Enter a unique ID used to identify the target. This will show in the **Target ID** column of the **Targets** tab. You can't change it later, so make sure you like it.

Configuration

Webhook URL: Add the full URL of your Slack Incoming Webhook. For example: `https://hooks.slack.com/services/T00000000/B00000000/XXXXXXXXXXXXXXXXXXXXXXXX`.

Retries

Honor Retry-After header: When toggled to `Yes` and the `retry-after` header is present, Cribl Edge honors any `retry-after` header that specifies a delay, up to a maximum of 20 seconds. Cribl Edge always ignores `retry-after` headers that specify a delay longer than 20 seconds.

Cribl Edge will log a warning message with the delay value retrieved from the `retry-after` header (converted to ms).

When toggled to `No` (the default), Cribl Edge ignores all `retry-after` headers.

Settings for failed HTTP requests: Automatically retries after unsuccessful response status codes, such as `429` (Too Many Requests) or `503` (Service Unavailable). Clicking **Add Setting** reveals a table where you can add retry parameters for individual failed HTTP requests. These include:

- **HTTP status code**: The individual status code for which the retry parameters apply.

- **Pre-backoff interval (ms)**: How long, in milliseconds, Cribl Edge should wait before initiating backoff. The maximum interval is 600,000 ms (10 minutes).
- **Backoff multiplier**: Base for exponential backoff. A value of `2` (default) means Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so forth.
- **Backoff limit (ms)**: The maximum backoff interval, in milliseconds, Cribl Edge should apply. Default (and minimum) is 10,000 ms (10 seconds); maximum is 180,000 ms (180 seconds).

Retry timed-out HTTP requests: When set to `Yes`, Cribl Edge to automatically retries HTTP requests that have timed out. Defaults to `No`. Enabling this option exposes the following settings:

- **Pre-backoff interval (ms)**: How long, in milliseconds, Cribl Edge should wait before initiating backoff. Maximum interval is `600,000 ms` (10 minutes).
- **Backoff multiplier**: Base for exponential backoff. A value of `2` (default) means Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so forth.
- **Backoff limit (ms)**: The maximum backoff interval, in milliseconds, Cribl Edge should apply. Default (and minimum) is 10,000 ms (10 seconds); maximum is 180,000 ms (180 seconds).

7.2.5.4. AWS SNS Notification Targets

You can send notifications to an Amazon Simple Notification Service (SNS) topic. This gives you access to a broad array of notification destinations, such as various AWS services, mobile push notifications, or text messages.

To add an Amazon SNS Notification target in Cribl Edge, go to **Manage** > **Notifications** > **Targets** > **Add Target**.

General Settings

Target ID: Enter a unique ID used to identify the target. This will show in the **Target ID** column of the **Targets** tab. It can't be changed later, so make sure you like it.

Configuration

Destination type: Defaults to **Topic ARN**. The SMS section below explains the **Phone number** option.

Region: Select the region associated with the Amazon S3 bucket.

Default Topic ARN: The default Amazon Resource Name (ARN) of the Amazon SNS topic to which you want to send notifications. Cribl Edge expects the ARN in a format like this:

`arn:aws:sns:region:account-id:MyTopic`.

If you use a non-AWS URL, the format must be:

`{url}/myQueueName` – for example, `https://host:port/myQueueName`.

Must be a JavaScript expression (which can evaluate to a constant value), enclosed in quotes or backticks. Can be evaluated only at initialization time. For example, if you're referencing a Global Variable: `https://host:port/myQueue-${C.vars.myVar}`. This value can be overridden by the notification event `__topicArn` field.

Phone number allowlist: A wildcard list of phone numbers that are allowed to receive SMS notifications. This is used when **Destination type** is set to **Phone number**.

Authentication

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Edge Nodes not in an AWS VPC, like those running a private cloud.

The **Manual** option exposes these additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

The values for **Access key** and **Secret key** can be a constant, or a JavaScript expression (such as `${C.env.MY_VAR}`) enclosed in quotes or backticks, which allows configuration with environment variables.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The Secret option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. To store a new, reusable secret, click **Create**.

Assume Role

Enable for SNS: Toggle to `Yes` to define an IAM Role to use, instead of automatically detecting one locally.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role. This is required only when assuming a role that requires this ID to delegate third-party access. For details, see AWS' documentation.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Post-Processing

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Maximum number of retries: The maximum number of retries before the output returns an error. The retries use an exponential backoff policy.

Endpoint: The SNS service endpoint. If empty, defaults to AWS' Region-specific endpoint. Otherwise, it must point to an SNS-compatible endpoint.

Signature version: Signature version to use for signing SNS requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (for example, self-signed certificates). Defaults to `Yes`.

SMS Notifications

You can use an Amazon SNS Notification target to send text messages (SMS) to a list of phone numbers. To do this, you'll need to set up a whitelist of phone numbers that are allowed to receive notifications.

1. Go to **Manage** > **Notifications** > **Targets** > **Add Target**.
2. Enter a unique **Target ID**.
3. Set the **Target type** to **AWS SNS**.
4. Set **Destination type** to **Phone number**.
5. Set **Region** to the region of the Amazon S3 bucket.
6. In **Default Phone number**, enter a comma-separated list of phone numbers that are allowed to receive notifications. This value can be overridden by the notification event `__phoneNumber` field. You can use `*` as the wildcard character.
   For example: `+15555550123, +15555551***`.
7. If desired, use **Phone number allowlist** to specify a wildcard list of allowed phone numbers.
8. Configure the remaining sections of the AWS SNS Notification target as described above.

9. Select **Save**.

Now, when you set up Notifications, you can select the new Amazon SNS target and specify any phone number that matches the configured whitelist.

7.2.5.5. Email Notification Targets

You can send notifications by email, using an SMTP server of your choice. Once you have configured the email notification target, you can specify the recipients and customize the subject line and message content. See Email Notifications for details on configuring an email Notification.

To add an email Notification target in Cribl Edge, go to **Manage** > **Notifications** > **Targets** > **Add Target**.

If you are a Cribl.Cloud user with an Enterprise license, you can use the default email Notification target. This target requires no configuration.

General Settings

Target ID: Enter a unique name to identify this email Notification target.

Configuration

Address: Identify the SMTP server by its hostname or IP address.

Port: Set the SMTP port. Use port `587` for SMTP Secure (SMTPS). You can also use port `25`. Use `465` when SSL/TLS is enabled. You can also use port `2525` if your email service provider supports this port as a backup when other ports are blocked by a network provider or a firewall.

Encryption type: Specify the encryption type used to secure SMTP communication. Options include:

- **STARTTLS**: Select this option to start the connection as plaintext, then upgrade it to a TLS-encrypted one if the server supports it. If the server doesn't support it, the connection remains plaintext.

  STARTTLS upgrades the connection to be encrypted but does not authenticate the server. Take additional steps to prevent man-in-the-middle attacks.

- **Require STARTTLS**: Select this option to require TLS. If the server doesn't support a secure connection, the connection will be dropped.
- **TLS (SMTPS)**: Select this option to use an encrypted connection from the start without requiring a subsequent connection upgrade.
- **None**: Select this option to use a plaintext connection.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path** or by another trusted CA (such as the system's CA).

From: Identify the email address of the sender.

Authentication

Username: The authentication principal (if required).

Password: The authentication credential (if required).

Testing Email Notification Targets

When you finish configuring your email Notification target, you can manage it, using the following options.

Click **Save & Test Target** to save your target and open a modal where you can enter an email to test your email Notification target. After configuration, the button text changes to **Test Target**.

Default Email Notification Target for Cribl.Cloud Enterprise Orgs

For Cribl.Cloud Enterprise orgs, a default email Notification target is available for Cribl Stream, Edge, and Search. You can use this target to route any email Notification to any valid email address.

This target appears in the **Notifications** > **Target** modal as `system_email`. You cannot modify or remove this target.

The `system_email` Notification target is managed by Cribl. It will be disabled in the unlikely event of abuse. If this target is disabled for a workspace, a log entry will appear in the Notifications Service logs. Select **Monitoring** > **Logs** > **Notifications Service** to view this log.

The `system_email` email Notification target is available only for Cribl.Cloud Enterprise organizations.

Sending Domain for Cribl.Cloud Email Notifications

Every Cribl.Cloud workspace and organization has a unique address for its email Notification target. Messages sent using this target will have a sender's address in the form `do-not-reply@<workspace>-<org>.criblcloud.email`. To ensure delivery, recipients should add the `criblcloud.email` domain to their email allowlists.

7.3. Licensing

Every Cribl Edge download package comes with a Free license that allows for processing of up to 1 TB/day. This license requires sending anonymized Telemetry Data to Cribl.

Enterprise, Standard, and Sales Trial licenses are entitled to a defined, per-license daily ingestion volume. These licenses do not require sending telemetry metadata.

> This page does not apply to Cribl.Cloud plans. For basic information about Cribl.Cloud Enterprise and Standard plans, see Cloud Pricing.

Managing Licenses – Adding and Renewing

On the Leader (or in a single-instance deployment), after submitting your newly untarred Cribl Edge software's registration page, you can add and manage licenses at **Settings** > **Global Settings** > **Licensing**. Click **Add License** to paste in a license key provided to you by Cribl.

This applies to Cribl Edge Standard and Enterprise licenses, which must be renewed annually. Free licenses are already onboard the download package, need not be added or managed here, and do not expire.

Airgapped Deployments

If your environment does not allow internet connectivity, you'll need to bypass the registration Web form when you deploy or upgrade Cribl Edge. To do so:

1. On the Leader's (or single instance's) host, navigate to `$CRIBL_HOME/local/cribl`.
2. Create a file named `license.yml`.
3. Paste in your license key in this format, then save the result:

```
licenses:
- <your-license-key-here>
```

Expiration/Renewal Notifications

In Cribl Edge 4.3 and later, the UI will prompt you to renew your paid license by displaying banner and drawer notifications, starting 30 days before license expiration. When you see these prompts, verify your license's expiration date and contact Cribl Sales at sales@cribl.io to renew.

Beyond these default notifications, you can configure custom License-Expiration Notifications. You can send these to PagerDuty, or to other external services via webhook.

Exemptions from License Quotas

Cribl does not require a separate license for sending data from Cribl Edge to Cribl Edge, such as sending from one Worker Group/Fleet managed by Leader Node A to a different Worker Group/Fleet managed by Leader Node B. In such situations, the same license used on Leader Node A can be used on Leader Node B.

Data generated by Cribl Internal Sources normally does not count against your ingestion quota.

If you are connecting Workers/Edge Nodes in a Cribl Stream Cloud hybrid deployment, you are granted additional exemptions to prevent double billing:

- Data transferred between Cribl Stream Workers via the Cribl HTTP and Cribl TCP Sources does not count against your ingestion quota.
- Data sent from Cribl Edge's Cribl HTTP and Cribl TCP Destinations to Cribl Stream is counted against quota only in Cribl Edge.
- Data generated by Datagens does not count against your ingestion quota.
- Dropped events do not count against your ingestion quota.

License Types

Cribl offers several Cribl Edge license types including Enterprise, Standard, and Free. For a current list of available license types and a detailed comparison of what's included in each, see Cribl Pricing.

Combining License Types

Multiple license types can coexist on an instance. However, only a **single type** of license can be effective at any one time. When multiple types coexist, the following method of resolution is used:

- If there are any unexpired Enterprise or Standard licenses – use only these licenses to compute the effective license.
- Else, if there are any Sales Trial licenses – use only Sales Trial licenses to compute the effective license.
- Else, if there exists a Free license – use only the Free license to compute the effective license.

License Expiration

When an Enterprise or Standard license expires, you can remove the old license to fall back to the Sales Trial or Free type. An expired Sales Trial license cannot fall back to a Free license.

To delete your expired license:

1. Select **Settings** then **Licensing**.
2. Locate your expired license, and click **Delete license**.
3. Restart the Leader.

Upon expiration of a paid license, if there is no fallback license, Cribl Edge will backpressure and block all incoming data.

Licensing in Distributed Deployments

With licenses that limit the number of Worker Processes/Edge Nodes, Cribl Edge will attempt to balance or rebalance Worker Processes (threads) as evenly as possible across all licensed Edge Nodes.

You configure licensing only on the Leader Node. (See Managing Licenses – Adding and Renewing.) The Leader will push license information down to Fleets as part of the heartbeat (Stream, Edge). There is no need to configure or store licenses directly on Edge Nodes.

Telemetry Data

A **Free** license requires sharing of telemetry **metadata** with Cribl. Cribl uses this metadata to help us understand how to improve the product and prioritize new features.

Telemetry payloads are sent from all Cribl Edge nodes (Leader and Workers), to an endpoint located at `https://cdn.cribl.io/telemetry/`.

Testing the Telemetry Endpoint's Connectivity

To manually test connectivity to the telemetry endpoint, especially if you are needing to configure a proxy, you can use the following command:

```
$ curl https://cdn.cribl.io/telemetry/
```

Expected response:

```
cribl /// living the stream!
```

If you get a 302 response code, check whether you've omitted the URL's trailing `/`.

Disabling Telemetry and Live Help

With an Enterprise or Standard license, you have the option to disable telemetry sharing from on-prem Cribl Edge. With a Free license, disabling telemetry will cause Cribl Edge to block inbound traffic within 24 hours.

If you would like an exception to disable telemetry in order to deploy in your environment, please contact Cribl Sales at sales@cribl.io, and we will work with you to issue licenses on a case-by-case basis.

Once you have received a license that removes the telemetry requirement, you can disable telemetry in Cribl Edge's UI at **Settings** > **Global Settings** > **System > General Settings > Upgrade & Share Settings > Share telemetry with Cribl**. Set the toggle to `No`.

Metadata Shared Through Telemetry

Your Cribl Edge instance shares metadata with Cribl per interval (roughly, every minute).

Details sent for nodes in any mode:

- Version
- OS Distribution/Version
- Kernel Version
- Instance's GUID
- License ID
- Earliest, Latest Time
- Number of Events In and Out, overall and by Source type and Destination type
- Number of Bytes In and Out, overall and by Source type and Destination type
- Number of Open, Closed, Active Connections
- Number of Routes
- Number of Pipelines
- Runtime Environment Indicators (AWS and Kubernetes)

Additional details sent for Leader nodes (when Cribl isn't sending managed node data):

- Runtime Environment
- OS Distribution / Version
- Version

- Kernel Version

How We Use Telemetry Data

With telemetry enabled, Cribl software sends usage data directly to Cribl. We securely store and encrypt this data on Cribl-managed servers, with restricted access.

Cribl fully anonymizes and aggregates this usage data in our systems of analysis. There, we use the aggregated metrics to improve Cribl products and services by analyzing circumstances around disruptions, opportunities for ingest efficiencies, and ways to optimize performance. Access is restricted to only those Cribl employees and contractors who require anonymized data to perform their jobs.

Cribl collects License IDs only to ensure that all data is being sent by a Cribl product. These IDs cannot be used to personally identify any user of Cribl software or Cribl cloud services.

For further details, see the Cribl Privacy Policy.

Licensing FAQ

How do I check my license type, restrictions, and/or expiration date?

Open Cribl Edge's **Settings** > **Global Settings** > **Licensing** page to see these details.

How can I track my actual data ingestion volume over the last 30 days?

Use the dashboard at **Monitoring** > **System** > **Licensing** to review your license consumption. For higher-fidelity metrics, you can enable the Cribl Internal Source's CriblMetrics option to forward metrics to your metrics Destination of choice, and run a report on `cribl.total.in_bytes`.

How does Cribl enforce license limits?

If your data throughput exceeds your license quota:

- Free and Standard licenses enforce data ingestion quotas through limits on the number of Worker Groups/Fleets and Worker/Edge Node Processes.
- Enterprise license keys turn off all enforcement.
- When an Enterprise or Standard license expires, Cribl Edge will attempt to fall back to a trial or free license, or – only if that fails – will block incoming data. For details, see Combining License Types.

I'm using LogStream 2.3.0 or higher, with its "permanent, Free" license. Why is LogStream claiming an expired license, and blocking inputs?

This can happen if you've upgraded from a LogStream version below 2.3.0, in which you previously entered this earlier version's Free (time-limited) license key. To remedy this, go to **Settings** > **Global Settings** > **Licensing**, click to select and expand your expired Free license, and then click **Delete license**. Cribl Edge will fall back to the new, permanent Free license behavior, and will restore throughput.

If I pull data from compressed S3 buckets, is my license quota applied to the compressed or the uncompressed size of the file objects?

To measure license consumption, Cribl Edge uses the uncompressed size.

Troubleshooting Resources

Cribl University offers a Troubleshooting Criblet on Switching from Free to Enterprise. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Troubleshooting Criblets and Advanced Troubleshooting short courses.

7.4. Uninstalling Cribl Edge From Linux

Removing a Cribl Edge installation, whether for a clean reinstall or permanently

---

Uninstalling a Single Edge Node

- (Optional) To prevent systemd from trying to start Cribl Edge at boot time, run the following command:
  `sudo $CRIBL_HOME/bin/cribl boot-start disable`
  If you're running both Cribl Edge and Cribl Stream on the same host, be sure to execute this command in the same mode (Edge or Stream) and the same directory in which you originally ran `cribl boot-start enable`.
- Stop Cribl Edge (stopping the main process).
- Back up any necessary configurations/data.
- Delete the `cribl` user: `userdel cribl`
- Delete the user group: `groupdel cribl`
- Remove the directory where Cribl Edge is installed.
- Directly delete any local files that Cribl Edge added while running – e.g., the `.dat` file and other local configs.

Uninstalling a Distributed Deployment

In a distributed deployment, repeat the above steps on:

- The Leader instance.
- All Edge Node instances.

7.5. Uninstalling Cribl Edge From Windows

You can uninstall Cribl Edge – whether for a clean reinstall or permanently – either via the Windows UI or via your command prompt.

Using the Windows UI

To uninstall Cribl Edge using Windows' graphical UI:

1. On your Windows Server, search for **Add or remove programs**.
2. In the **Apps & features** list, scroll to **Cribl Edge**, and click **Uninstall**.
3. In the confirmation dialog, click **Yes**.

A success message will confirm removal.

This uninstall process does not automatically delete the `local`, `log`, `pid`, and `state` directories, because those folders include contents generated after the original installation completed. To completely remove the application, you must explicitly delete the folders from the filesystem.

Using the Command Prompt

If you installed Cribl Edge using the Msi Installer, you can uninstall it by running the following at a command prompt:

```
- msiexec /x cribl-<version>-<build>-<arch>.msi
```

7.6. Fleet Settings

A Fleet's General Settings allows you to configure teleporting, throughput throttling, logging, upgrading, and security at the Fleet level.

To edit a Fleet's General Settings:

1. Select the Fleet, then click **Fleet Settings**.
2. Configure the following settings per Fleet.

General Settings

Fleet Configuration

You can configure the following options on this tab:

Description: Optionally, add or edit a description of the Fleet's purpose.

Tags: Use tags to organize Fleets into logical categories. Then, you can search by tags on the **View all Groups** (Stream) or **View all Fleets** (Edge) interface. This search filters the list of Fleets, showing only those with the tag you entered.

Enable teleporting to Workers: Use this toggle to enable or disable authenticated access to Workers' UI from the Leader (Stream, Edge).

In Cribl Edge 4.1.2 and later, you can configure the following option in Shutdown Settings.

API Server Settings

General

You can set the following options for the API server. In **General Settings**, open **API Server Settings** and select **General**.

Host: The hostname or IP address you want to bind the API server to. Defaults to `0.0.0.0`.

Port: API port to listen to. Defaults to `9000`.

TLS

For information on TLS options, see the documentation for any Source or Destination that supports TLS.

Advanced

See Authentication Controls.

Default TLS Settings

See the Securing and Monitoring topic.

Limits

The **Limits** tab provides access to controls for metrics, storage, metadata, jobs, the Redis cache and connections to it, and CPU settings.

Metrics

See Controlling Metrics Volume.

Storage

You can configure the following options for storage. In **General Settings**, open **Limits** and then select **Storage**:

Max sample size: Maximum file size, in binary units (`KB`, `MB`), for sample data files. Maximum: `3 MB`. Default: `256 KB`.

Min free disk space: The minimum amount of disk space on the host before various features take measures to prevent disk usage (KB, MB, etc.). Default: `5 GB`.

Max PQ size per Worker Process: Highest accepted value for the **Max queue size** option used in individual Sources' and Destinations' persistent queues. Default: `1 TB`. Consult Cribl Support before increasing beyond this value.

Metadata

Event metadata sources: List of event metadata sources to enable. No sources are enabled by default.

Jobs

Disable jobs/tasks: Set to `Yes` by default in Edge Fleets and `No` by default in Stream Groups. In Edge, Nodes no longer poll the Leader for upgrade jobs. Setting this to `Yes` in Edge reduces application load from the Leader.

Nodes running 4.4.4 and older still upgrade via jobs and will honor the Jobs settings, even with the **Disable jobs/tasks** toggle enabled.

Job Limits

Disable Jobs/Tasks: When enabled, the Edge Nodes won't poll the Leader for jobs/tasks. The job limits settings below will not affect Edge Nodes on version 4.5.0 and newer. Edge Nodes running 4.4.4 and older still use these jobs settings even if jobs/tasks are disabled here.

Concurrent Job Limit: The total number of jobs that can run concurrently. Defaults to `10`.

Concurrent System Job Limit: The total number of **system** jobs that can run concurrently. Defaults to `10`. Minimum `1`.

Concurrent Scheduled Job Limit: The total number of **scheduled** jobs that can run concurrently. This limit is set as an offset relative to the **Concurrent Job Limit**. Defaults to `-2`.

Skipped jobs indicate that a Group's **Concurrent Job Limit** has been reached or exceeded. Increase this limit to reduce the number of skippable jobs. For resource-intensive jobs, this might require deploying more Worker Nodes.

Task Limits

Concurrent Task Limit: The total number of tasks that a Worker Process can run concurrently. Defaults to `2`. Minimum `1`.

Concurrent System Task Limit: The number of system tasks that a Worker Process can run concurrently. Defaults to `1`. Minimum `1`.

Max Task Usage Percentage: Value, between `0` and `1`, representing the percentage of total tasks on a Worker Process that any single job may consume. Defaults to `0.5` (i.e., 50%).

Task Poll Timeout: The number of milliseconds that a Worker's task handler will wait to receive a task, before retrying a request for a task. Defaults to `60000` (i.e., 60 seconds). Minimum `10000` (10 seconds).x

Completion Limits

Artifact Reaper Period: Interval on which Cribl Edge attempts to reap jobs' stale disk artifacts. Defaults to `30m`.

Finished Job Artifacts Limit: Maximum number of finished job artifacts to keep on disk. Defaults to `100`. Minimum `0`.

Finished Task Artifacts Limit: Maximum number of finished task artifacts to keep on disk, per job, on each Worker Node. Defaults to `500`. Minimum `0`.

Task Manifest and Buffering Limits

Manifest Flush Period: The rate (in milliseconds) at which a job's task manifest should be refreshed. Defaults to `100` ms. Minimum `100`, maximum `10000`.

Manifest Max Buffer Size: The maximum number of tasks that the task manifest can hold in memory before flushing to disk. Defaults to `1000`. Minimum `100`, maximum `10000`.

Manifest Reader Buffer Size: The number of bytes that the task manifest reader should pull from disk. Defaults to `4kb`.

Job Dispatching: The method by which tasks are assigned to Worker Processes. Defaults to `Least In-Flight Tasks`, to optimize available capacity. `Round Robin` is also available.

Job Timeout: Maximum time a job is allowed to run. Defaults to `0`, for unlimited time. Units are seconds if not specified. Sample entries: `30`, `45s`, `15m`.

Task Heartbeat Period: The heartbeat period (in seconds) for tasks to report back to the Leader/API. Defaults to `60` seconds. Minimum `60`.

Redis

Cache

Key TTL in seconds: Maximum time to live of a key in the cache (seconds). `0` indicates no limit. Defaults to `10 minutes`.

Max # of keys: Maximum number of keys to retain in the cache. `0` indicates no limit. Defaults to `0`.

Max cache size (bytes): Maximum number of bytes to retain in the cache. `0` indicates no limit. Defaults to `0`.

Service period (seconds): Frequency of cache limit enforcement. Defaults to every `30 seconds`.

Server assisted: Defaults to `No`. When toggled to `Yes`, the following control appears.

Client tracking mechanism: Mechanism for invalidation message delivery. In default mode, the server remembers which keys a client has requested and only sends invalidations for those, using more Redis server memory. In broadcast mode, it sends all invalidations, requiring more processing by Cribl Edge.

Connections

Reuse Redis connections: Toggle on if you want Cribl Edge to try to reuse Redis connections when multiple Redis Functions (or references to them) are present. When enabled, displays the following additional control:

- **Max number of connections**: The maximum number of identical connections allowed before Cribl Edge tries to reuse connections. Defaults to `0`, meaning unlimited connections are allowed (equivalent to leaving **Reuse Redis connections** toggled off). Setting a non-zero integer value forces Cribl Edge to try to reuse connections for each individual Worker Process (**not** to reuse connections **among** Worker Processes).

To understand why and when to employ these controls, see Reusing Redis Connections.

Other

CPU profile TTL: The time-to-live for collected CPU profiles.

Default managed node heartbeat period: How many seconds a managed Node will wait to send back a heartbeat to the Cribl control plane.

Proxy Settings

Use proxy env vars: Honors the `HTTP_PROXY/HTTPS_PROXY` environment variables. Defaults to `Yes`.

Sockets

Directory: Holds sockets for inter-process communication (IPC), such as communications between a load-balancing process and a Worker Process. Defaults to `/tmp` (your system's temp directory).

Shutdown Settings

Drain timeout (sec): Determines how long a Cribl server will wait for writes to complete before the server shuts down on individual Worker Processes. If you notice that Workers are under-ingesting available data

upon shutdown or restart, increase the `10`–second default. Acceptable range of values: minimum `1` second, maximum `600` seconds (10 minutes).

Worker Processes

For details about this left tab's **Process count**, **Minimum process count**, and **Memory (MB)** controls, see Sizing and Scaling.

Process count and **Minimum process count** are not configurable on Cribl Edge, where each Edge Node is automatically allocated `1` Worker Process.

The following controls are also available on this tab to optimize Worker Processes' throughput on startup.

Max connections at startup: Maximum number of connections accepted at Worker Process startup. Defaults to `1`. Enter a negative integer for unlimited connections.

Startup throttling duration (ms): Maximum time (in milliseconds) to continue throttling connections after Worker Process startup. Defaults to `10000` ms (10 sec.) Enter `0` to disable throttling.

Load throttle %: Sets a threshold to prevent overwhelming Workers. If 90% of a Worker Process' CPU utilization readings exceed this threshold over one minute, the process will reject new connections until the CPU load stabilizes. Another process that is below the threshold will accept the connection the next time it is established. Defaults to `0`% (no throttling). Enter a percentage between `1`–`100` to enable throttling.

You can configure the CPU saturation threshold, but the 90% sampling trigger is not configurable. Also, `_raw stats > cpuPerc` values might diverge from your **Load throttle %** threshold. This is because `cpuPerc` is sampled and averaged once per minute, whereas the **Load throttle %** is evaluated every second, with a rolling 1-minute lookback sample. (These intervals are also not configurable.)

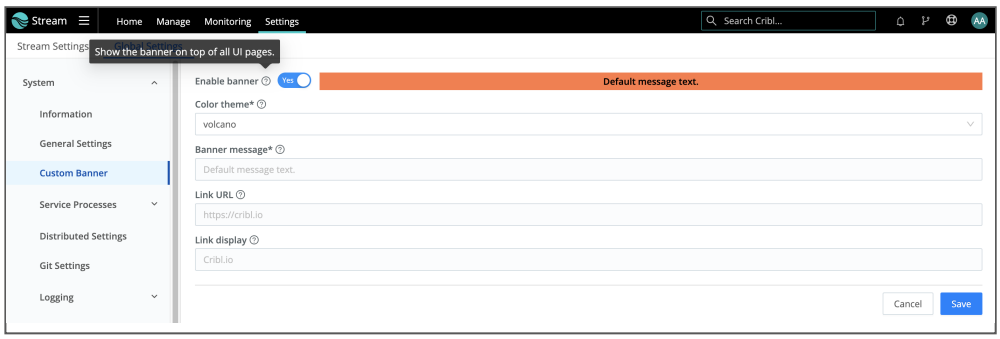
Other Settings

This page's remaining options work essentially the same way as their **Global Settings** counterparts. Use the following links for details about: logging levels/redactions, access management security, scripts, and diagnostics.

7.7. Custom Banners

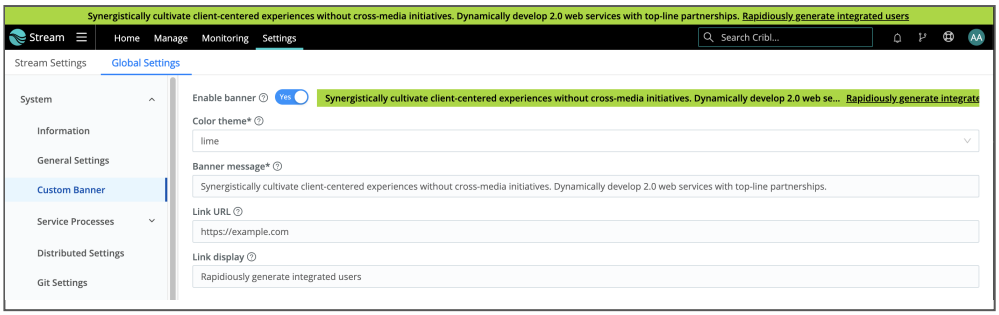
In version 4.3 and later, Cribl admins can define a custom banner from the Leader. This will be displayed at the top of your Cribl Organization's UI, above all Cribl apps. You can display one banner at a time.

Your banner can display warnings, requirements, neutral updates, or good news about new features or options. It can contain a mixture of text, a raw URL, or a URL with a friendly/short label. The banner does not appear unless you enable and configure it, as outlined below.



Configuring/enabling a custom banner

1. From Cribl Edge's top nav, select **Settings** > **Global Settings**.
2. From the resulting left nav, select **System** > **Custom Banner**.
3. Toggle the **Enable banner** slider on. This exposes the configuration options below.
4. From the **Color theme** drop-down, select a background color matching your banner's severity. (Cribl reserves dark red, light blue, and light green for system notifications, so the drop-down offers alternative shades of these colors.)
5. In the **Banner message** field, enter a single-line message of up to 1,024 characters. (The displayed banner will truncate characters that cannot fit across one line of a given viewport.)
6. In the **Link URL** field, optionally enter a hyperlink to append to the message.
7. In the **Link display** field, optionally give your **Link URL** a text label. This can include up to 512 characters, but a long label can force-truncate the **Banner message**. A short label can help everything fit.
8. Click **Save** to display your banner.
9. Commit your changes from the Leader.



Custom banner, configured and displayed

To hide the banner, toggle the **Enable banner** slider off, and again **Save** and commit your changes. Your banner's configuration will remain intact, ready to re-enable and/or revise when you're ready.

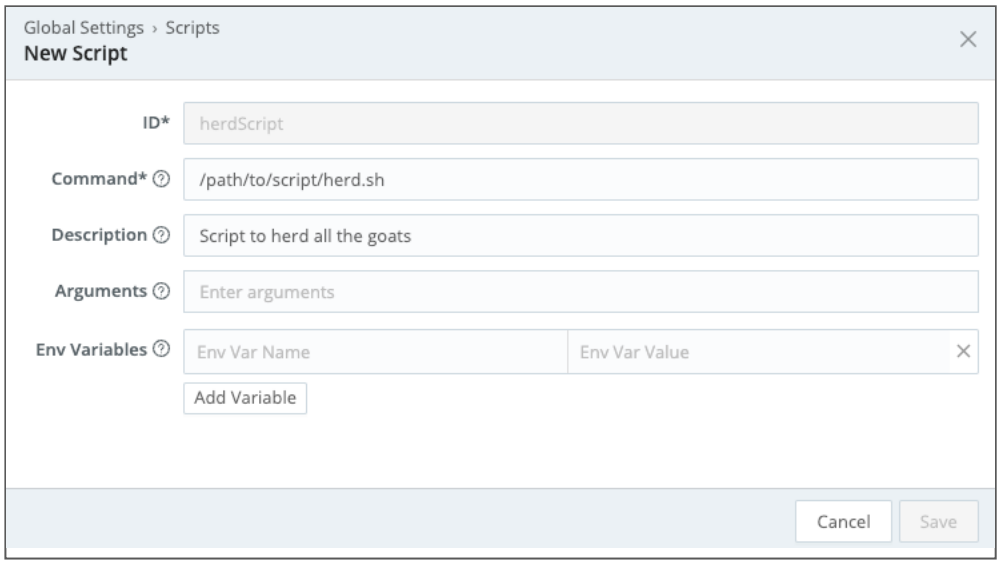
7.8. Scripts

Admins can run scripts (for example, shell scripts) from within Cribl Edge by configuring and executing them at **Settings** > **Global Settings** > **Scripts**. Scripts are typically used to call custom automation jobs or, more generally, to trigger tasks on demand. For example, you can use Scripts to run an Ansible job, or to place a call to another automation system, when Cribl Stream configs are updated.

With Great Power Comes Great Responsibility!

Scripts will allow you to execute almost anything on the system where Cribl Edge is running. Make sure you understand the impact of what you're executing before you do so!

To add a new script, click **Add script** on the scripts page.



Settings > Scripts page

The **New Script** modal provides the following fields:

- **ID**: Unique ID for this script.
- **Command**: Command to execute for this script.
- **Description**: Brief description about this script. Optional.
- **Arguments**: Arguments to pass when executing this script.
- **Env variables**: Extra environment variables to set when executing script.

Scripts in Distributed Deployments

- Scripts can be deployed from Leader Node, but can be **run** only locally from each Worker Node.

- If the Script command is referencing a file (for example, `herd.sh`), that file must exist on the Cribl Stream instance. In other words, the Script management interface cannot be used to upload or manage script files.

8. Access Management

Cribl Edge provides a range of access-management features for users with different security requirements.

Where Can I Find Access Control Details?

See the following topics, according to your needs:

- Authentication: Authenticating users via local basic auth or external options (SSO, Splunk, LDAP).
- Members and Permissions: Available in Cribl Edge 4.2 and later. Fine-grained access control configurable at separate levels (Organization, product, Fleet, and lower-level resources like Stream Projects and Search datasets).
- Local Users: Cribl Edge's original Role-based model for creating users, and for managing their access across a Cribl deployment.
- Roles: Cribl Edge's original RBAC model for managing Roles and Policies, and for assigning them to users.

Prerequisites (Restrictions on Restrictions)

Permission- and Role-based access control can be enabled only on distributed deployments (Stream, Edge) with an Enterprise license. With other license types and/or single-instance deployments (Stream, Edge), note that **all users will have full administrative privileges.**

Which Access Method Should I Use?

Cribl currently supports both the new Members/Permissions and the legacy Users/Roles models, and these models are cross-compatible for many use cases. However, certain purposes require you to choose a specific model:

- **Cribl.Cloud** now relies only on Members/Permissions. See Cribl.Cloud Organization-level Permissions starting at Inviting Members, and product- and lower-level Permissions starting at Product-Level Permissions.

  Cribl.Cloud's Organization-level Permissions include an Owner superuser. This option currently has no counterpart at the on-prem (customer-managed) Organization level.

- **Stream Projects** and Subscriptions, in Cribl Stream 4.2 and later, rely only on Members/Permissions. See Project-Level Permissions.

- **GitOps** integration authorization requires the legacy `gitops` Role. This legacy Role currently has no counterpart Permission.
- **Collectors**: The `collect_all` Role specifically enables creating, configuring, and running Collection jobs on all Stream Worker Groups. This legacy Role currently has no counterpart Permission.
- **Notifications**: The `notification_admin` Role specifically enables creating and receiving all Notifications. This legacy Role currently has no counterpart Permission.
- **Sources**, **Destinations**, **Pipelines**, and **Routes** are examples of other lower-level resources (below the product level) that can be shared with Local Users only by configuring custom access in legacy `policies.yml` configuration files.

  > Customizing these files is currently supported only with on-prem (customer-managed) deployments, not on Cribl.Cloud.

- **Search granular resources** (datasets, dataset providers, and search results) can be shared via Members/Permissions. For details, see the Search Sharing topic.

8.1. Authentication

User authentication in Cribl Edge

Cribl Edge supports **local**, **Splunk**, **LDAP**, **SSO/OpenID Connect**, and **SSO/SAML** authentication methods, depending on license type.

Local Authentication

To set up local authentication, navigate to **Settings** > [**Global Settings** >] **Access Management** > **Authentication** and select **Local**.

You can then manage users through the **Settings** > [**Global Settings** >] **Access Management** > **Local Users** UI. All changes made to users are persisted in a file located at `$CRIBL_HOME/local/cribl/auth/users.json`.

For Windows Edge Nodes, the changes are persisted in a file located at `C:\ProgramData\Cribl\local\cribl\auth\users.json`.

This is the line format, and note that both usernames and passwords are case-sensitive:

```
{"username":"user","first":"Goat","last":"McGoat","disabled":"false",
"passwd":"Yrt0MOD1w8OzyMYB8WMcEleOtYESMwZw2qIZyTvueOE"}
```

The file is monitored for modifications every 60s, and will be reloaded if changes are detected.

Adding users through direct modification of the file is also supported, but not recommended.

If you edit `users.json`, maintain each JSON element as a single line. Otherwise, the file will not reload properly.

Manual Password Replacement

Manual Password Replacement is for on-prem deployments only. For Cribl Edge in Cribl.Cloud, Cribl recommends setting up Fallback Access; if you do need to reset your password, you must contact Cribl.

To manually add, change, or restore a password, replace the affected user's `passwd` key-value pair with a `password` key, in this format: `"password":"<newPlaintext>"`. Cribl Edge will hash all plaintext password(s), identified by the `password` key, during the next file reload, and will rename the plaintext `password` key.

Starting with the same `users.json` line above:

```
{"username":"user","first":"Goat","last":"McGoat","disabled":"false",
"passwd":"Yrt0MOD1w8OzyMYB8WMcEleOtYESMwZw2qIZyTvueOE"}
```

...you'd modify the final key-value pair to something like:

```
{"username":"user","first":"Goat","last":"McGoat","disabled":"false",
"password":"V3ry53CuR&pW9"}
```

Within at most one minute after you save the file, Cribl Edge will rename the `password` key back to `passwd`, and will hash its value, re-creating something resembling the original example.

Set Worker/Edge Node Passwords

In a distributed deployment (Edge, Stream), once a Worker/Edge Node has been set to point to the Leader Node, Cribl Edge will set each Worker/Edge Node's admin password with a randomized password that is different from the admin user's password on the Leader Node. This is by design, as a security precaution. But it might lead to situations where administrators cannot log into a Worker/Edge Node directly, and must rely on accessing them via the Leader.

To explicitly apply a known/new password to your Edge Node, you set and push a new password to the Fleet. Here's how, in the Leader Node's UI:

1. From the top nav, select **Manage**.
2. Select the desired Fleet.
3. From the Fleet's submenu, select **Group Settings**.
4. Select **Local Users**, then expand the desired user.
5. Update the **Password** and **Confirm Password** fields and select **Save**.

Every 10 seconds, the Worker/Edge Node will request an update of configuration from the Leader, and any new password settings will be included.

> On Cribl.Cloud, Group Settings for hybrid Workers include the above **Local Users** options, but Groups for Cribl-managed Workers do not. For alternatives, see the Cribl.Cloud Launch Guide and Cribl.Cloud SSO Setup docs.

Authentication Controls

You can customize authentication behavior at **Settings** > [**Global Settings** >] **General Settings** > **API Server Settings** > **Advanced.** The options here include:

- **Logout on Roles change**: If role-based access control is enabled, determines whether users are automatically logged out of Cribl Edge when their assigned Roles change. Defaults to `Yes`.
- **Auth-token TTL**: Sets authentication tokens' valid lifetime, in seconds. Defaults to `3600` (60 minutes = 1 hour); must be at least `1` sec.
- **Session idle time limit**: Sets how long (in seconds) Cribl Edge will observe no user interaction before invalidating user's session tokens. Defaults to `3600` (60 minutes = 1 hour); must be at least `60` sec.
- **Login rate limit**: Sets the number of login attempts allowed over a (selectable) unit of time. Defaults to `2/second`.
- **HTTP header**: Enables you to specify one or more custom HTTP headers to be sent with every response.

Token Renewal and Session Timeout

Here is how Cribl Edge sets tokens' valid lifetime by applying the **Auth-token TTL** field's value:

- When a user logs in, Cribl Edge returns a token whose expiration time is set to {login time + **Auth-token TTL** value}.
- If the user is idle (no UI activity) for the configured token lifetime, they are logged out.
- As long as the user is interacting with Cribl Edge's UI in their browser, Cribl Edge continually renews the token, resetting the idle-session time limit back by the **Auth-token TTL** value.

The cribl.secret File

When Cribl Edge first starts, it creates a `$CRIBL_HOME/local/cribl/auth/cribl.secret` file. This file contains a key that is used to generate auth tokens for users, encrypt their passwords, and encrypt encryption keys.

Default local credentials are: `admin/admin`

Back up and secure access to this file by applying strict permissions – e.g., `600`.

External Authentication

While configuring any external auth method, make sure you don't get locked out of Cribl Edge! Enable the **Fallback on fatal error** or **Allow local auth** toggle until you're certain that external auth is working as intended. If you do get locked out, refer back to Manual Password Replacement for the remedy.

All external auth methods require either an Enterprise or a Standard license. They're not supported with a Free license.

Cribl Edge Roles and role mapping are supported **only** with an Enterprise license. With a Standard license, all your external users will be imported to Cribl Edge in the `admin` Role.

This topic covers the following external authentication providers:

- Splunk Authentication
- LDAP Authentication

Authentication for Single Sign-On

Each authentication method that Cribl Edge supports for on-prem SSO has its own separate topic:

- SSO/OpenID Connect Authentication
- SSO/SAML Authentication

For deployments in Cribl.Cloud, see Cribl.Cloud SSO Setup.

Splunk Authentication

Splunk authentication is very helpful when deploying Cribl Edge in the same environment as Splunk. This option requires the user to have Splunk `admin` role permissions. To set up Splunk authentication:

Navigate to **Settings** > [**Global Settings** >] **Access Management > Authentication > Type** and select **Splunk**. This exposes the following controls.

- **Host**: Splunk hostname (typically a search head).
- **Port**: Splunk management port (defaults to `8089`).
- **SSL**: Whether SSL is enabled on Splunk instance that provides authentication. Defaults to `Yes`.
- **Fallback on fatal error**: Attempt local authentication if Splunk authentication is unsuccessful. Defaults to `No`. If toggled to `Yes`, Cribl Edge will attempt local auth only **after** a failed Splunk auth. Selecting `Yes` also exposes this additional option:

- **Fallback on bad login**: Attempt local authentication if the supplied user/password fails to log in on Splunk. This similarly defaults to `No`.

The Splunk search head does not need to be locally installed on the Cribl Edge instance.

For distributed deployments with an Enterprise License, the next (and final) step is to configure Role Mapping settings.

LDAP Authentication

You can set up LDAP authentication as follows:

Navigate to **Settings** > [**Global Settings** >] **Access Management** > **Authentication** > **Type**, and select **LDAP**. This exposes the following controls.

- **Secure**: Enable to use a secure LDAP connections (`ldaps://`). Disable for an insecure (`ldap://`) connection.
- **LDAP servers**: List of LDAP servers. Each entry should contain `host:port` (e.g., `localhost:389`).
- **Bind DN**: Distinguished Name of entity to authenticate with LDAP server. E.g., `'cn=admin,dc=example,dc=org'`.
- **Password**: Distinguished Name password used to authenticate with LDAP server.
- **User search base**: Starting point to search LDAP for users, e.g., `'dc=example,dc=org'`.
- **Username field**: LDAP user search field, e.g., `cn` or `(cn (or uid)`. For Microsoft Active Directory, use `sAMAccountName` here.
- **User search filter**: LDAP search filter to apply when authenticating users. Optional, but recommended: If empty, all users in your domain will be able to log in with the default Cribl Role. This example would enable only users in the `StreamAdmins` group to log in – they would then be mapped to any Roles specified in Role Mapping Settings:
  `(&(objectClass=user)(memberof=CN=StreamAdmins,OU=Groups,DC=cribl,DC=io))`
  This simpler filter example would enable anyone in a group called admin," who does **not** have a department attribute that starts with "123," to log in:
  `(&(group=admin)(!(department=123*)))`
- **Group search base**,
  Group member field,
  Group membership attribute,
  Group search filter,
  Group name field: These settings are used only for LDAP authentication with role-based access control. See Role-Based LDAP Authentication, below.
- **Connection timeout (ms)**: Defaults to `5000`.

- **Reject unauthorized**: Valid for secure LDAP connections. Set to `Yes` to reject unauthorized server certificates.
- **Fallback on fatal error**: Attempt local authentication if LDAP authentication is down or misconfigured. Defaults to `No`. If toggled to `Yes`, local auth will be attempted only **after** a failed LDAP auth. Selecting `Yes` also exposes this additional option:
    - **Fallback on bad login**: Attempt local authentication if the supplied user/password fails to log in on the LDAP provider. Defaults to `No`.

For distributed deployments with an Enterprise License, the next (and final) step is to configure Role Mapping settings.

Role-Based LDAP Authentication

When configuring LDAP authentication with role-based access control (RBAC), you **must** use the following settings to import user groups. (The UI does not enforce filling these fields. When using LDAP without roles, ignore them.)

- **Group search base**: Starting point to search LDAP for groups, e.g., `dc=example,dc=org`.
- **Group member field**: LDAP group search field, e.g., `member`.
- **Group membership attribute**: Attribute name of LDAP user object. Determines group member attribute's value, which defines group's allowed users. This field accepts `dn`, `cn`, `uid`, or `uidNumber`. If none of these are specified, falls back to `dn`.
- **Group search filter**: LDAP search filter to apply when retrieving groups for authorization and mapping to Roles., e.g., `(&(cn=cribl*)(objectclass=group))`.
- **Group name field**: Attribute used in objects' DNs that represents the group name, e.g., `cn`. Cribl Edge does not directly read this attribute from group objects; rather, it must be present in your groups' DN values. Match the attribute name's original case (upper, lower, or mixed) when you specify it in this field. In particular, Microsoft Active Directory requires all-uppercase group names (e.g., `CN`).

For distributed deployments with an Enterprise License, the next (and final) step is to configure Role Mapping settings.

Role Mapping Settings

This section is displayed only on **distributed** deployments (Edge, Stream) with an Enterprise License. For details on mapping your external identity provider's configured groups to corresponding Cribl Edge user access Roles, see External Groups and Roles. The controls here are:

- **Default Role**: Default Cribl Edge Role to assign to all groups not explicitly mapped to a Role.

- **Mapping**: On each mapping row, enter an external group name (case-sensitive) on the left, and select the corresponding Cribl Edge Role in the right drop-down list. Click **Add Mapping** to add more rows.

8.2. Members And Permissions

Define and manage fine-grained access control across Cribl products and resources

---

In Cribl Edge 4.2 and later, Members and Permissions are available as the successors to Cribl's original role-based access control (RBAC) model of Local Users and Roles/Policies. The earlier model is still supported across most of the Cribl product suite. However, Cribl.Cloud invitations and Stream Projects have fully transitioned to the new model.

Permission-based access control is available only on distributed deployments (Stream, Edge) with an Enterprise license or Cribl.Cloud plan. With other license/plan types and/or single-instance deployments, all users will have full administrative privileges.

In the current release, existing Local Users display incorrectly on **Settings** > **Members** pages with `No Access` Permissions. This is a display-only bug: These users' original Roles still function as configured. For details and fix timeline, please see Known Issues.

When you first deploy Cribl Edge with the above prerequisites, you will be granted the Organization-level `Admin` Permission. Using this Permission, you can then assign additional Permissions to yourself and other Members, as outlined below in Organizations and Assigning Group-/Fleet-Level Permissions.

Why Move My Cheese?

Members and Permissions enable the finer-grained access control and authorization that many of our users have requested. You can now assign different Cribl users access and capabilities independently at multiple levels:

- Organization.
- Product: Stream, Edge, or Search.
- Group/Fleet.
- Resource: Stream Project, Search Dataset or Dataset Provider, etc.

The benefits?:

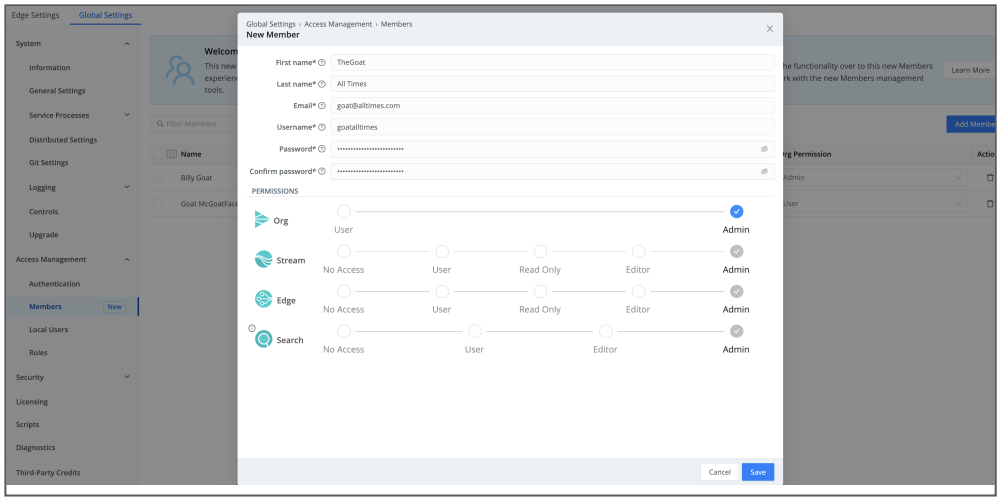
- Greater security in protecting access to resources and configuration.
- Greater flexibility in authorizing users on different parts of your Cribl suite (i.e., Organization).
- Ability to segment access and authorization to relevant teams.
- Ability to support a greater diversity of users, given these options.

Cribl plans to eventually retire its earlier RBAC model. But for now, both models are supported, to enable users to experiment and switch over at their own pace. As outlined below, several components of the new versus old model are currently interchangeable.

Organizations

Cribl.Cloud has, from the start, provided the concept of an Organization as a container for the deployment of a whole suite of Cribl products (Stream, Edge, and Cloud-only Search). The Members and Permissions model brings this concept to customer-managed (on-prem) deployments. Here, to access the Organization-level Members UI as an Admin:

1. From Cribl Edge's top nav, select **Settings** > **Global Settings**.
2. From the left nav, select **Members**.
3. Here, you can click an existing Member's row to change their access Permissions or other details, or click **Add Member** to grant a new user access to your Cribl Organization.



Organization > Members/Permissions UI

For Cribl.Cloud's updated counterpart to this on-prem UI for managing Organization-level Permissions, see Inviting Members.

Members and Local Users

The top half of the **Members** configuration modal is almost identical to Cribl's original Local Users modal. In the current release, Members and Local Users are interchangeable: Members whom you add here can be reconfigured in the traditional **Local Users** UI (using legacy Roles instead of Permissions), and anyone you add in Local Users will also show up here in **Members**.

So let's focus on the the modal's bottom half, which is more interesting – the new graphical **Permissions** management.

Organization- and Product-Level Permissions

Each newly created Member starts out with **User** Permission at the Organization level, and with the self-explanatory **No Access** Permission on each product. You can use the toggles to assign the following Permission levels.

For Cribl.Cloud Organization-level Permissions, see Cribl.Cloud docs' Member Permissions section.

Organization-Level Permissions

Permission	Description
User	The most basic Permission. At the Organization level, gets the Member into the system, without conferring any access to peer Members, products, or lower-level resources. (Admins can freely assign these Members varying Permissions at lower levels.)
Admin	This is a superuser Permission. At the Organization level, allows creating, viewing, updating, and deleting all Members. (Automatically inherits Product-level `Admin` Permissions and resource-level `Maintainer` Permissions, which confer comparably broad capabilities.)

For Stream and Edge, the Product-level Permissions below apply to both Cribl.Cloud and on-prem deployments. For Cribl Search's more-specific Product-level Permissions, see Search Member Permissions.

Product-Level Permissions (Stream and Edge)

Permission	Description
User	The most basic Permission. At the Product level, makes the Member assignable to Fleets and resources, with no initial access to any.
Read Only	At the Product level, this Permission is designed specifically to enable support personnel to help other users by viewing (but not modifying) their configurations. Allows viewing all Members, Fleets, Settings, Leader commits, and legacy Local Users and Roles, with no configuration capabilities. (A `Read Only` Permission at the Product level automatically inherits `Read Only` Permissions on all Fleets and lower-level resources.)

Permission	Description
Editor	Allows viewing all Groups and Monitoring pages. (Configuring these options requires `Admin`-level permission on the product and/or Fleet.) Automatically inherits the `Editor` Permission on Fleets, and the `Maintainer` Permission on lower-level resources.
Admin	This is a superuser Permission at the Product level. It allows creating, viewing, updating, and deleting all Fleets and resources; managing Fleet Mappings; adding, updating, and restarting Edge Nodes; and managing Notifications and Notification targets. (Automatically inherited from the Organization-level `Admin` Permission in on-prem deployments, and from the Org-level `Owner` Permission in Cribl.Cloud deployments. Automatically inherits `Admin` Permissions on all Fleets and `Maintainer` Permissions on lower-level resources.)

Once Members are in your Organization, you can assign or reassign the above Organization- and Product-level Permissions. Return to **Settings** > **Global Settings** > **Members** and click the Member's row to reopen the same configuration modal.

Group- and Fleet-Level Permissions

Cribl Edge Permissions at the Fleet level generally mirror those at the Product level:

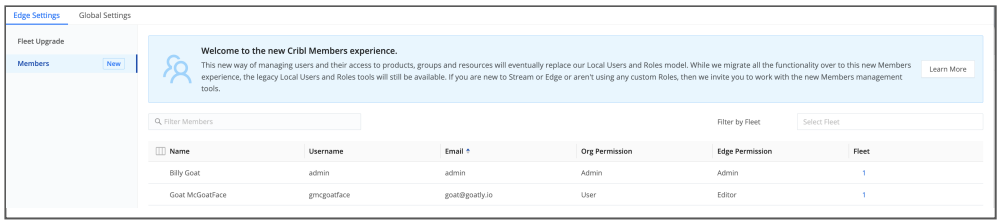
Permission	Description
User	The most basic Permission. Gets the Member into the Fleet, where Admins can then assign them access on individual resources. (Unless Members have a higher Permission at the Product level, they have no initial access to Fleets or resources.)
Read Only	Designed for support personnel. Allows viewing all Fleet-level Settings, encryption keys, certificates, secrets, scripts, Sources, Destinations, Pipelines, Packs, Routes, QuickConnect connections, Knowledge objects, Notifications and Notification targets, Edge Subfleets and their Settings, and Stream Projects and Subscriptions, with no configuration capabilities. (Automatically inherited from a Product-level `Read Only` Permission. Automatically inherits `Read Only` Permissions on lower-level resources.)
Editor	Allows creating, viewing, updating, and deleting all Fleet-level encryption keys, certificates, secrets, scripts, Sources, Destinations, Pipelines, Packs, Routes, QuickConnect connections, Knowledge objects, and Notifications and Notification targets. Allows committing configuration changes, but not deploying them. (An `Editor` Permission at the Product level automatically inherits `Editor` Permissions on all Fleets.)
Admin	Superuser Permission. Allows creating, viewing, updating, and deleting all Fleet-level access management (Members' Permissions), Settings, encryption keys, key management system (KMS) settings, certificates, secrets, scripts, Sources, Destinations, Pipelines, Packs, Routes, QuickConnect connections, Knowledge objects, Notifications and Notification targets, and Stream Projects/Subscriptions.

Permission	Description
	Allows adding, updating, and restarting Edge Nodes. Allows committing and deploying configuration changes. On Edge, provides CRUD capabilities on Subfleets' Settings and access management identical to those on the parent Fleet. (Automatically inherited from the Product-level `Admin` Permission. Automatically inherits `Maintainer` Permissions on lower-level resources.)

Assigning Group-/Fleet-Level Permissions

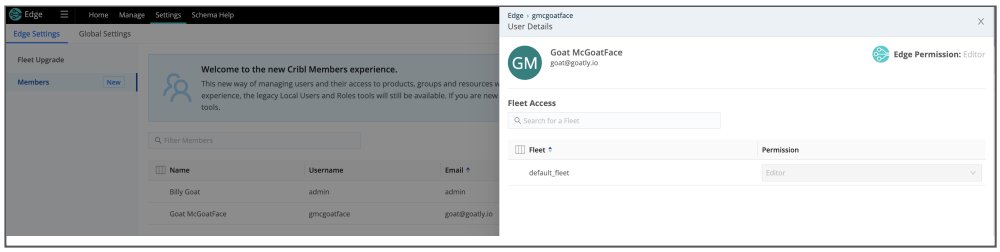
To assign Permissions on a Fleet:

1. Select **Settings** > **Global Settings** > **Members**.
2. Make sure this Member has an **Org Permission** level sufficient to support the Permission you want to assign on the Fleet.
3. From the top-left product switcher, select **Cribl Edge**. Then select (depending on the product) **Settings** > **Stream Settings** > **Members** or **Settings** > **Edge Settings** > **Members**.



Edge Settings > Members

4. Click each applicable Member's row to open their **User Details** drawer.



Product Members > User Details

5. Grant the Member the Permission level you want on this Fleet.

To change Members' existing Permissions, repeat the same steps above.

Resource-Level Permissions

Certain Cribl products provide Permission-based access to particular resources.

Stream Projects

In Cribl Stream 4.2 and later, Stream Projects and Subscriptions rely entirely on the Members/Permissions access control first introduced in Stream 4.2. These enable you to assign different users different levels of access on individual Projects. The 4.1.x `project_user` Role and `ProjectSourceSubscribe` Policy, which provided access to all Projects, are now retired.

This is a breaking change. If you are upgrading with Projects configured in v.4.1.x, those Projects' users will be visible in Stream as Members, and you'll need to assign them new Permissions.

For details, see Adding Users to Projects.

Search Resources

For Permissions on Search resources, see the Search docs' Sharing topic.

Other Resources

Access to certain resources can be managed only via legacy Local Users and Roles.

GitOps integration: Authorization to enable and manage GitOps requires the legacy `gitops` Role. This legacy Role currently has no counterpart Permission.

Collectors: The `collect_all` legacy Role specifically enables creating, configuring, and running Collection jobs on all Stream Worker Groups. This Role currently has no counterpart Permission.

Notifications: The `notification_admin` legacy Role specifically enables creating and receiving all Notifications. This legacy Role currently has no counterpart Permission.

Sources, Destinations, Pipelines, and Routes are examples of other lower-level resources that can be shared with Local Users only by configuring custom access in legacy `policies.yml` configuration files.

Customizing these files is currently supported only with on-prem (customer-managed) deployments, not on Cribl.Cloud.

Inheritance

In the current release, Members' Permissions at certain levels determine the Permissions that Admins can assign them at lower levels. Here is the current inheritance scheme:

- Members with the `Admin` Permission at the Organization level will be locked to the `Admin` Permission on Products, to the `Editor` Permission on Fleets, and to the `Maintainer` Permission on lower-level resources.
- Members with the Product-level `Editor` Permission will be locked to the `Editor` Permission on Fleets, and to the `Maintainer` Permission on lower-level resources.
- Members with the `User` Permission at each level can be assigned varying Permissions at the next level down. **`User` is the most malleable Permission.**
- On Stream Projects, the `Maintainer` Permission is available **only** to Members with higher-level `Admin` or `Editor` Permissions. (They're automatically assigned this Permission via inheritance.)
- Search resources are more flexible: Here, the `Maintainer` Permission can be shared with Members who have the `User` Permission at the Product level.
- Members with the Product-level `Read Only` Permission will be locked to the `Read Only` Permission on Fleets and on lower-level resources.

UX Customization

Cribl Edge's UI will be presented differently to users who are assigned Permissions that impose access restrictions. Some controls will be visible but disabled, and some internal-search and log results will be limited, depending on each user's Permissions.

Access to the same objects via Cribl Edge's API and CLI will be similarly filtered, with appropriate error reporting. E.g., if a user tries to commit and deploy changes on a Worker Group/Fleet where they are not authorized, they might receive a CLI error message like this: `git commit-deploy command failed with err: Forbidden`

Legacy Roles and Policies

To facilitate a smooth transition to Cribl's new access-control model – and to provide backward compatibility for customers still using our earlier Roles/Policies model – our pre-4.2 Roles have been supplemented with new counterparts to most of the new Permissions listed above.

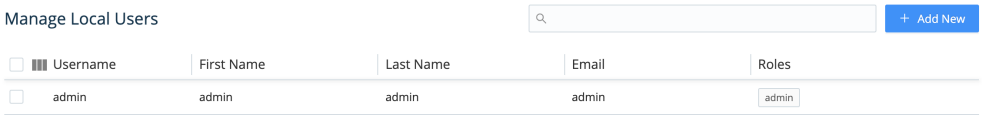
8.3. Local Users

This page covers how to create and manage Cribl Edge users, including their credentials and (where enabled) their access roles. These options apply if you're using the **Local** Authentication type, which is detailed here.

Creating and Managing Local Users

On the Leader Node, you manage users by selecting **Settings** > **Global Settings** > **Access Management** > **Local Users**. In single-instance deployments (Stream, Edge), you select **Settings** > **Access Management** > **Local Users**.

The resulting **Local Users** page will initially show only the default `admin` user. You are operating as this user.

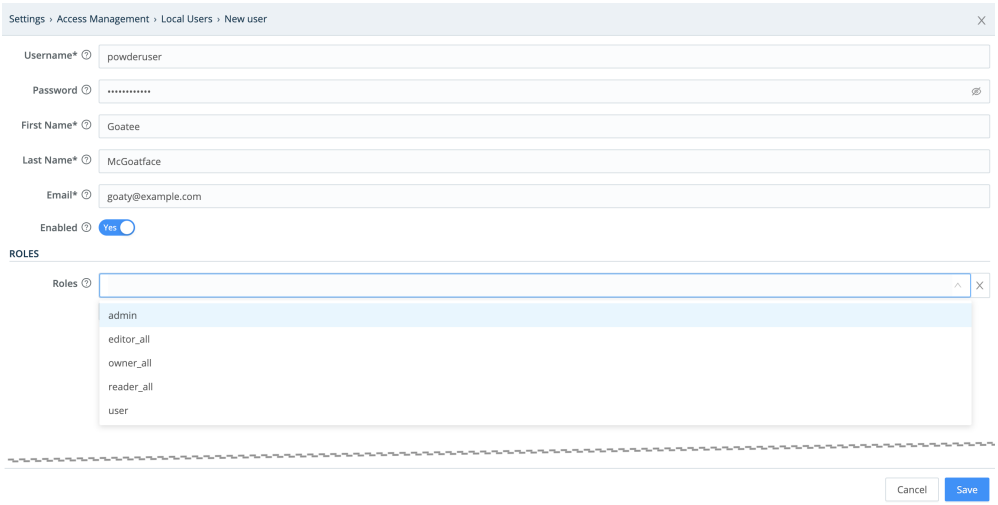


Managing users

To create a new Cribl Edge user, click **New User**. To edit an existing user, click anywhere on its row. With either selection, you will see the modal shown below.

The first few fields are self-explanatory: they establish the user's credentials. Usernames and passwords are case-sensitive.

If you choose to establish or maintain a user's credentials on Cribl Edge, but prevent them from currently logging in, you can toggle **Enabled** to `No`.



Entering and saving a user's credentials

Logged-in users can change their own Cribl Edge password from the **Local Users** page, by clicking on their own row to open a **Local Users** modal that manages their identity.

Password Rules

All passwords must:

- Contain eight or more characters.
- Use characters from three or more of the following categories:
  - Lowercase letters.
  - Uppercase letters located after the first character in the password.
  - Digits located before the last character in the password.
  - Non-alphanumeric ASCII characters such as `#`, `!`, or `?`.
  - Non-ASCII characters such as `ñ`, `€`, or emoji.

As of Cribl Edge version 4.5.0, these rules apply for all passwords, whether or not Cribl Edge is running in FIPS mode, with one exception:

- For Cribl Edge not running in FIPS mode, local users whose passwords existed before Cribl Edge 4.4.4 was released, can continue to use their passwords, even if those passwords do not satisfy the rules.
- When these users change to a new password, the new password must satisfy the rules.
- New users must create passwords that satisfy the rules.

If you get locked out of your account, you need to reset your password manually.

Adding Roles

If you've enabled role-based access control you can use the modal's bottom **Roles** section to assign access Roles to this new or existing user.

For details, see Roles. Role-based access control can be enabled only on distributed deployments (Edge, Stream) with an Enterprise license. With other license types and/or single-instance deployments (Edge, Stream), all users will have full administrative privileges.

Click **Add Role** to assign each desired role to this user. The options on the **Roles** drop-down reflect the Roles you've configured at **Settings** > **Global Settings** > **Access Management** > **Roles**.

Note that when you assign multiple Roles to a user, the Roles' permissions are additive: This user is granted a superset of the highest permissions contained in all the assigned Roles.

When you've configured (or reconfigured) this user as desired, click **Save**.

By default, Cribl Edge will log out a user upon a change in their assigned Roles. You can defeat this behavior at **Settings** > **Global Settings** > **General Settings > API Server Settings > Advanced > Logout on roles change.**

8.4. Roles

Define and manage access-control Roles and Policies

Cribl Edge offers role-based access control (RBAC) to serve these common enterprise goals:

- **Security**: Limit the blast radius of inadvertent or intentional errors, by restricting each user's actions to their needed scope within the application.
- **Accountability**: Ensure compliance, by restricting read and write access to sensitive data.
- **Operational efficiency**: Match enterprise workflows, by delegating access over subsets of objects/resources to appropriate users and teams.

Role-based access control is available only on distributed deployments (Stream, Edge) with an Enterprise license or Cribl.Cloud plan. With other license/plan types and/or single-instance deployments, all users will have full administrative privileges.

Roles, Meet Permissions

In Cribl Edge 4.2 and later, the Roles/Policies model described on this page exists in parallel with a new, more flexible Members/Permissions model, which will eventually replace it. To provide cross-compatibility, we have added several new Default Roles and Default Policies that are counterparts to new Permissions.

As noted below, these cross-compatible Roles/Policies support customers who still choose to configure Local Users with Roles and Policies. Your configured Local Users appear interchangeably in the new Members UI, and vice versa.

Known Issue

In the current release, existing Local Users display incorrectly with `No Access` Permissions on **Settings** > **Members** pages. This is a display-only bug: These users' Roles still function as originally configured. For details and fix timeline, please see Known Issues.

RBAC Concepts

Cribl Edge's RBAC mechanism is designed around the following concepts, which you manage in the UI:

- **Roles**: Logical entities that are associated with one or multiple **Policies** (groups of permissions). You use each Role to consistently apply these permissions to multiple Cribl Edge **users**.
- **Policies**: A set of **permissions**. A Role that is granted a given Policy can access, or perform an action on, a specified Cribl Edge object or objects.
- **Permissions**: Access rights to navigate to, view, change, or delete specified **objects** in Cribl Edge.
- **Users**: You map Roles to Cribl Edge users in the same way that you map **user groups** to users in LDAP and other common access-control frameworks.

Users are independent Cribl Edge objects that you can configure even without RBAC enabled. For details, see Local Users.

How Cribl Edge RBAC Works

Cribl Edge RBAC is designed to grant arbitrary permissions over objects, attributes, and actions at arbitrary levels.

In Cribl Edge v.2.4.x through 4.1.x, Roles are customizable only down to the Worker Group/Fleet level. E.g., you can grant Edit permission on Worker Group/Fleet `WG1` to User A and User B; but you cannot grant them finer-grained permissions on child objects such as Pipelines, Routes, etc. through the UI.

Setting these granular permissions requires creating custom Policies in `policies.yml` configuration files. Note that this option is currently supported only with on-prem (customer-managed) deployments, not on Cribl.Cloud.

Cribl Edge's UI will be presented differently to users who are assigned Roles that impose access restrictions. Controls will be visible but disabled, and search and log results will be limited, depending on each user's permissions.

Access to the same objects via Cribl Edge's API and CLI will be similarly filtered, with appropriate error reporting. E.g., if a user tries to commit and deploy changes on a Worker Group/Fleet where they are not authorized, they might receive a CLI error message like this: `git commit-deploy command failed with err: Forbidden`

Cribl Edge Roles can be integrated with external authorization/IAM mechanisms, such as LDAP and OIDC and mapped to their respective groups, tags, etc.

Using Roles

Cribl Edge ships with a set of default Roles, which you can supplement.

Default Roles

These Roles ship with Cribl Edge by default.

Organization-Level Roles

Note that some of the Roles below have no counterpart Permission in Cribl's newer Members/Permissions model.

Name	Description	Permission Equivalent
admin	Superuser – authorized to do anything in the system.	Organization Admin
gitops	Ability to sync the Cribl Edge deployment to a remote Git repository.	N/A
notification_admin	Read/write access to all Notifications.	N/A
user	Default Role that gets only a home/landing page to authenticate. This is a fallback for users who have not yet been assigned a higher Role by an admin.	Organization User
project_user	Read/write access to the simplified Stream Projects UI and related data Subscriptions. **Deprecated as of v.4.2.** Instead assign **Editor**, as the most comparable new Project-level Permission. The more-permissive **Maintainer**, and the more-restrictive **Read Only**, are also available Permissions.	Project Editor

Stream Roles

Name	Description	Permission Equivalent
stream_user	Basic Role for Stream.	Stream User.
stream_reader	Allows viewing all Members, Worker Groups, Settings, Leader commits, and legacy Local Users and Roles, with no configuration capabilities.	Stream Read Only.

Name	Description	Permission Equivalent
stream_editor	Allows viewing all Groups and Monitoring pages.	Stream Editor.
stream_admin	Superuser Role at the Product level	Stream Admin.

Edge Roles

Name	Description	Permission Equivalent
edge_user	Basic Role for Edge.	Edge User.
edge_reader	Allows viewing all Members, Fleets, Settings, Leader commits, and legacy Local Users and Roles, with no configuration capabilities.	Edge Read Only.
edge_editor	Allows viewing all Groups and Monitoring pages.	Edge Editor.
edge_admin	Superuser Role at the Product level.	Edge Admin.

Fleet–Level Roles

Name	Description	Permission Equivalent
owner_all	Read/write access to (and Deploy permission on) all Fleets.	N/A
editor_all	Read/write access to all Fleets.	N/A
reader_all	Read-only access to all Fleets.	N/A
collect_all	Ability to create, configure, and run Collection jobs on all Worker Groups.	N/A

Search Roles

Name	Description	Permission Equivalent
search_user	Basic Role for Search.	Search **User**.
search_editor	Manage datasets, dataset providers, dashboards and settings.	Search **Editor**.
search_admin	Superuser Role at the Product level.	Search **Admin**.

Cribl **strongly recommends** that you do not edit or delete these default Roles. However, you can readily clone them (see the **Clone Role** button in the next section's screenshots), and modify the duplicates to meet your needs.

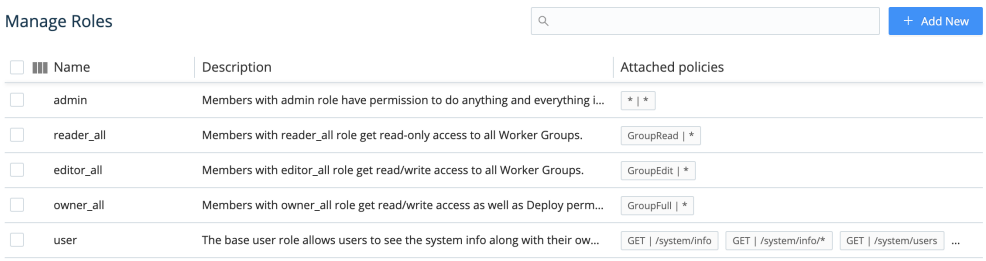
Initial Installation or Upgrade

When you first install Cribl Edge with the prerequisites to enable RBAC (Enterprise license and distributed deployment), you will be granted the **admin** Role. Using this Role, you can then define and apply additional Roles for other users.

You will similarly be granted the **admin** Role upon upgrading an existing Cribl Edge installation from pre-2.4 versions to v. 2.4 or higher. This maintains backwards-compatible access to everything your organization has configured under the previous Cribl Edge version's single Role.

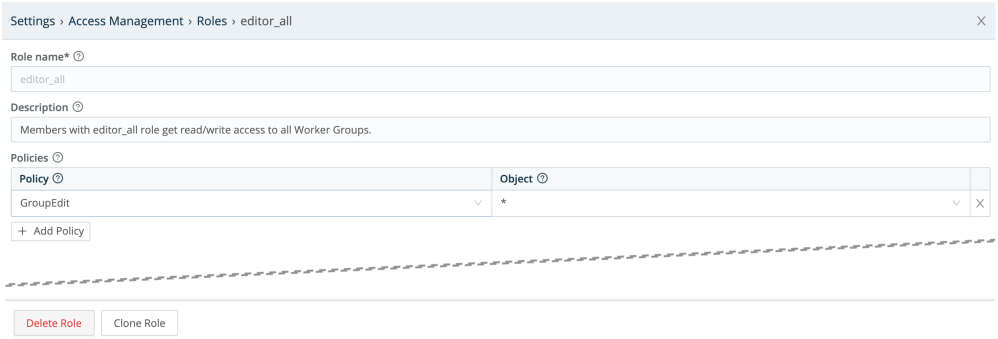
Adding and Modifying Roles

In a distributed environment, you manage Roles at the Leader level, for the entire deployment. On the Leader Node, select **Settings** > **Global Settings** > **Access Management** > **Roles**.



Manage Roles page

To add a new Role, click **New Role** at the upper right. To edit an existing Role, click anywhere on its row. Here again, either way, the resulting modal offers basically the same options.



Add/edit Role modal

The options at the modal's top and bottom are nearly self-explanatory:

Role name: Unique name for this Role. Cannot contain spaces.

Description: Optional free-text description.

Delete Role: And…it's gone. (But first, there's a confirmation prompt. Also, you cannot delete a Role assigned to an active user.)

Clone Role: Opens a **New role** version of the modal, duplicating the **Description** and **Policies** of the Role you started with.

The modal's central **Policies** section (described below) is its real working area.

Adding and Removing Policies

The **Policies** section is an expandable table. In each row, you select a Policy using the left drop-down, and apply that Policy to objects (i.e., assign permissions on those objects) using the right drop-down.

Let's highlight an example from the above screen capture of Cribl Edges built-in Roles: The `editor_all` Role has the `GroupEdit` Policy, with permission to exercise it on any and all Worker Groups/Fleets (as indicated by the `*` wildcard).

Policies on the left, objects on the right

To add a new Policy to a Role:

1. Click **Add Policy** to add a new row to the **Policies** table.
2. Select a Policy from the left column drop-down.
3. Accept the default object on the right; or select one from the drop-down.

To modify an already-assigned Policy, just edit its row's drop-downs in the **Policies** table.

To remove a Policy from the Role, click its close box at right.

In all cases, click **Save** to confirm your changes and close the modal.

Default Policies

In the **Policies** table's left column, the drop-down offers the following default Policies:

Fleet-Level Policies

Name	Description	Permission Equivalent
`GroupRead`	The most basic Fleet-level permission. Enables users to view a Fleet and its configuration, but not modify or delete the config.	Fleet-level Read Only.
`GroupEdit`	Building on `GroupRead`, grants the ability to also change and commit a Fleet's configuration.	Fleet-level Editor.
`GroupFull`	Building on `GroupEdit`, grants the ability to also deploy a Fleet.	Fleet-level Admin.
`GroupCollect`	Grants the ability to create, configure, and run Collectors on a Worker Group.	N/A
`GroupUser`	Access Fleet.	Fleet-level User.

Project-Level Policies

Name	Description	Permission Equivalent
`ProjectMaintain`	Grants ability to edit or delete the Project and its settings.	Project-level Maintainer.
`ProjectRead`	Can configure connections among the Project's Subscriptions, Packs, and Destinations, but cannot modify or delete these resources.	Project-level Read Only.
`ProjectEdit`	Can view Project and Subscription settings and connections, but not modify or delete them.	Project-level Editor.

Product-Level Policies

Name	Description	Permission Equivalent
`ProductUser`	Makes the Member assignable to Worker Groups and resources, with no initial access to any.	Product-level User.
`LimitedProductUser`	Similar to `ProductUser`, but omits the ability to read or act on all the endpoints within a Fleet.	N/A
`ProductAdmin`	Superuser Permission at the Product level.	Product-level Admin.

Search Policies

Name	Description	Permission Equivalent
`DatasetMaintain`	Full access to dataset configuration.	Search datasets ◉Maintainer.
`DatasetRead`	Can view and search, but not modify dataset configurations.	Search datasets ◉Read Only.
`DatasetProviderMaintain`	Full access to dataset provider configuration.	Search dataset providers ◉Maintainer.
`DatasetProviderRead`	Can view and search, but not modify dataset provider configurations.	Search dataset providers ◉Read Only.
`SearchUser`	Search data and view results shared with the user	Search Product-level ◉User.
`SearchMaintainer`	Manage datasets, dataset providers, dashboards and settings.	Search Product-level ◉Editor.
`DashboardMaintain`	Can edit and delete a Search dashboard.	Search dashboard ◉Maintainer.
`DashboardRead`	Can view and use a Search dashboard.	Search dashboard ◉Read Only.

General Policies

Name	Description	Permission Equivalent
`*` (wildcard)	Grants **all** permissions on associated objects.	N/A

Internal Policies

The following policies are internal building blocks for Product-specific Policies. Do not add them directly to Roles.

Name	Description	Permission Equivalent
`Product`	N/A	N/A
`BaseProductUser`	N/A	N/A
`MaintainBase`	N/A	N/A
`SearchBase`	Similar to `SearchUser`: Can view Search resources shared with the Member.	N/A

Use Policies As-Is

By design, the default Policies that ship with Cribl Edge cannot be modified via the UI or API. Do not attempt to modify them by other means. Breaking the built-in model could undermine your intended access-control protections, opening your deployment and data to security vulnerabilities.

Objects and Permissions

In the **Policies** table's right column, use the drop-down to select the Cribl Edge objects on which the left column's Policy will apply. (Remember that in v. 2.4, the objects available for selection are specific Worker Groups/Fleets, or a wildcard representing all Worker Groups/Fleets.) For example:

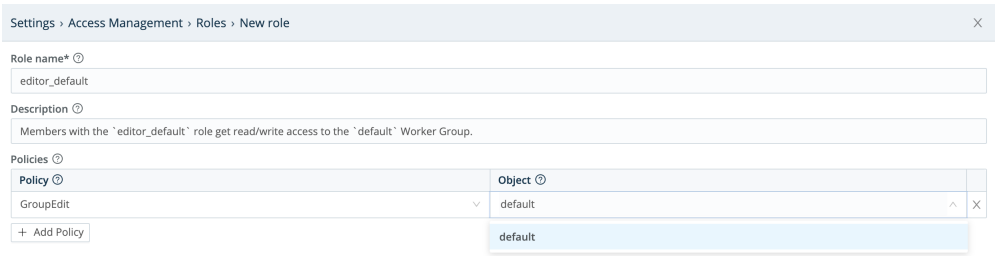
- `Worker Group <id>`
- `NewGroup2`
- `default` (Worker Group)
- `*` (all Worker Groups)

Extending Default Roles

Here's a basic example that ties together the above concepts and facilities. It demonstrates how to add a Role whose permissions are restricted to a particular Worker Group/Fleet.

Here, we've cloned the `editor_all` Role that we unpacked above. We've named the clone `editor_default`.

We've kept the `GroupEdit` Policy from `editor_all`. But in the right column, we're restricting its object permissions to the `default` Worker Group/Fleet that ships with Cribl Edge.



Cloning a default Role

You can readily adapt this example to create a Role that has permissions on an arbitrarily named Worker Group/Fleet of your own.

Roles and Users

Once you've defined a Role, you can associate it with Cribl Edge users. On the Leader Node, select **Settings** > **Global Settings** > **Access Management** > **Local Users**. For details, see Local Users.

Note that when you assign multiple Roles to a given user, the Roles' permissions are additive: This user is granted a superset of all the permissions contained in all the assigned Roles.

By default, Cribl Edge will log out a user upon a change in their assigned Roles. You can defeat this behavior at **Settings** > **Global Settings** > **General Settings > API Server Settings > Advanced > Logout on roles change.**

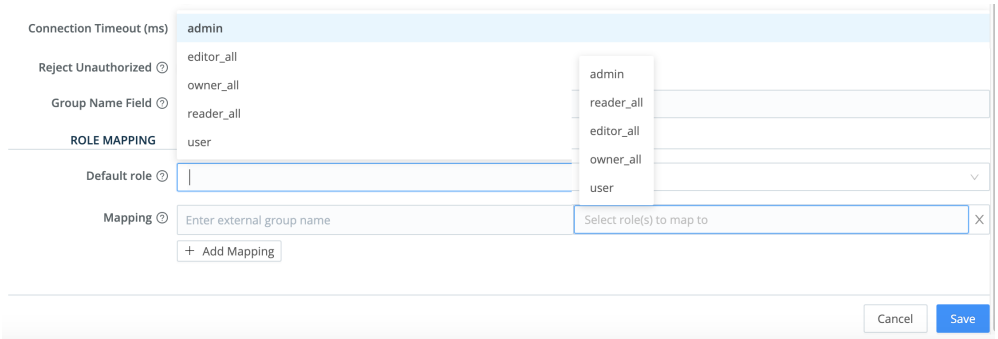
External Groups and Cribl Edge Roles

You can map user groups from external identity providers (LDAP, Splunk, or OIDC) to Cribl Edge Roles, as follows:

1. Select **Settings** > **Global Settings** > **Access Management** > **Authentication**.
2. From the **Type** drop-down, select **LDAP**, **Splunk**, or **OpenID Connect**, according to your needs.

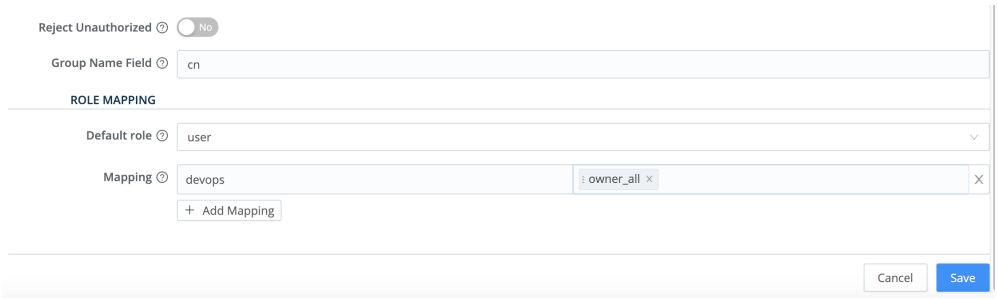
3. On the resulting **Authentication Settings** page, configure your identity provider's connection and other basics. (For configuration details, see the appropriate Authentication section.)
4. Under **Role Mapping**, first select a Cribl Edge **Default role** to apply to external user groups that have no explicit Cribl Edge mapping defined below.
5. Next, map external groups as you've configured them in your external identity provider (left field below) to Cribl Edge Roles (right drop-down list below).
6. To map more user groups, click **Add Mapping**.
7. When your configuration is complete, click **Save**.

Here's a composite showing the built-in Roles available on both the **Default role** and the **Mapping** drop-downs:



Mapping external user groups to Cribl Edge Roles

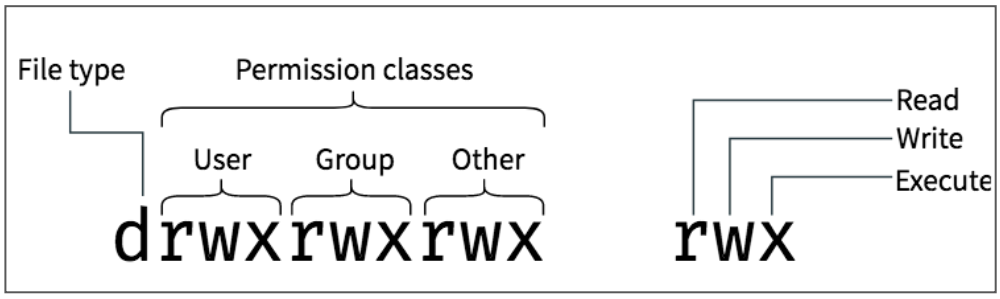
And here, we've set a conservative **Default Role** and one explicit **Mapping**:



External user groups mapped to Cribl Edge Roles

8.5. Using ACLs To Allow Cribl Edge To Read Files

Running Cribl Edge as an unprivileged user is a best practice. However, without modifying the default Linux permissions, you will run into issues in accessing files owned by other users.



Default Linux permissions

Linux systems allow you to layer an Access Control List (ACL) on top of the default Linux permission set. With ACLs, you can apply a more specific set of permissions to a file or directory without (necessarily) changing the base ownership and permissions. For details, see Introduction to ACLs.

As an example, you might want to read data from the `/var/log` directory. This directory is typically owned by the `root` user, with a permission set of `750` on the directory. This means the `cribl` user will not be able to read or list the files in the directory, because the `Other` group has zero permissions.

To achieve compliance with benchmarks such as CIS or NIST, we can use the ACLs to grant the `cribl` group access to this folder and any files, without disturbing the current permissions.

CIS Benchmark 4.2.3

Make sure that permissions are configured on all log files. Log files must have the correct permissions to ensure that sensitive data is archived and protected. `Other/world` should not have the ability to view this information. `Group` should not have the ability to modify this information.

To accomplish this, we can grant the `cribl` group `read` and `execute` access to the files and directories inside `/var/log`, by running this command:

```
setfacl -Rm g:cribl:r-X /var/log
```

Breaking down the command's options':

- `-R` : Recursive

- `-m` : Modify
- `g:cribl` = `cribl` group, this could be `u:cribl` if you wanted to limit to the `cribl` user.
- `r-X`: Read and execute. Capital `X` means execute only on directories.

This modifies only the current files in the directory, if you want the appropriate ACL applied to any future files created here, add the `-d` flag (for default):

```
setfacl -Rdm g:cribl:r-X /var/log
```

Now, any rotated or created files will apply the ACL set.

Checking the ACLs

To verify the ACLs on a file or directory, run the following command:

```
getfacl <file or folder>
```

This will output a listing of the applied ACLs, including the directory's defaults:

```
$ getfacl /<directory>
# file: <file>
# owner: <owner>
# group: <group>
user::rwx
group::rwx
other::---
default:user::rwx
default:user:<user>:rwx
default:group::rwx
default:mask::rwx
default:other::---
```

Installing ACL Utilities

The ACL utilities might not be installed, by default, on the OS. For example, on Ubuntu (Debian-based) systems, you will need to install the `acl` package. For Debian-based tools using apt:

```
apt install acl
```

For Red Hat-based tools using yum:

```
yum install acl
```

8.6. SSO With Okta And OIDC

Cribl Edge supports SSO/OIDC user authentication (login/password) and authorization (user's group membership, which you can map to Cribl Roles). Using OIDC will change the default `Log in` button on the login page to a button labeled `Log in with OIDC` which redirects to the configured Identity Provider (IDP).

If you are a Cribl Edge admin and want to offer single sign-on (SSO) to your users, you can choose OpenID Connect (OIDC) as the authentication type, then configure an IDP to use OIDC within its single sign-on flow. Once configuration is complete (several steps later), the Cribl Edge login page will send users to the IDP's login UI. Besides the IDP, some settings will refer to the Service Provider (SP), which in this context is your Cribl Edge instance.

The IDP can be Okta or Google, among others. Configuring SSO requires going back and forth between Cribl Edge and the IDP's UI. In this page, we walk through the process for configuring SSO with OIDC, using Okta as the IDP.

The last part of the walkthrough is a complete description of the @{product **Authentication** settings (with OIDC selected as the authentication method). Skip directly to this section if you just need a UI reference, or if your IDP is not Okta. For non-Okta deployments, please join us on Cribl's Community Slack at https://cribl-community.slack.com/ and share your questions.

Make sure you don't get locked out of Cribl Edge! Enable the **Allow local auth** toggle until you're certain that external auth is working as intended. If you do get locked out, refer to Manual Password Replacement for the remedy.

Like other external auth methods, OpenID Connect requires either an Enterprise or a Standard license. It is not supported with a Free license.

Plan Your Mapping of Okta Groups to Cribl Edge Roles

In Okta, admins organize their users in groups. In Cribl Edge, there are no user groups, but there are Roles. Your task includes **mapping** Okta groups to Cribl Edge Roles.

- Mapping groups to Roles is possible only for Cribl Edge deployments that are in Distributed mode, with an Enterprise license applied.
- If you are running Cribl Edge in Single-instance mode, you cannot map Okta groups to Cribl Edge Roles, although you can still set up SSO with Okta.

As you think through how best to map your Okta groups to Cribl Edge Roles, keep these principles in mind:

- An Okta group can map to more than one Cribl Edge Role.
- A Cribl Edge Role can map to more than one Okta group.
- If a user has multiple Roles, Cribl Edge applies the union of the most permissive levels of access.
- Cribl Edge automatically assigns the `default` Role to any user who has no mapped Roles.

The example below illustrates how multiple mappings work: The groups in Mapping **b** and **c** each map to multiple Roles, while both the `reader_all` and `editor_cloud` Roles map to multiple groups.

Mapping	Okta Group	Cribl Edge Role(s)
a.	Cribl Admins	`admin`
b.	Cloud Admins	`reader_all`, `editor_cloud`
c.	Security Team	`reader_all`, `editor_cloud`, `editor_firewall`

Cribl Edge Roles and role mapping are supported **only** with an Enterprise license. With a Standard license, all your external users will be imported to Cribl Edge in the `admin` role.

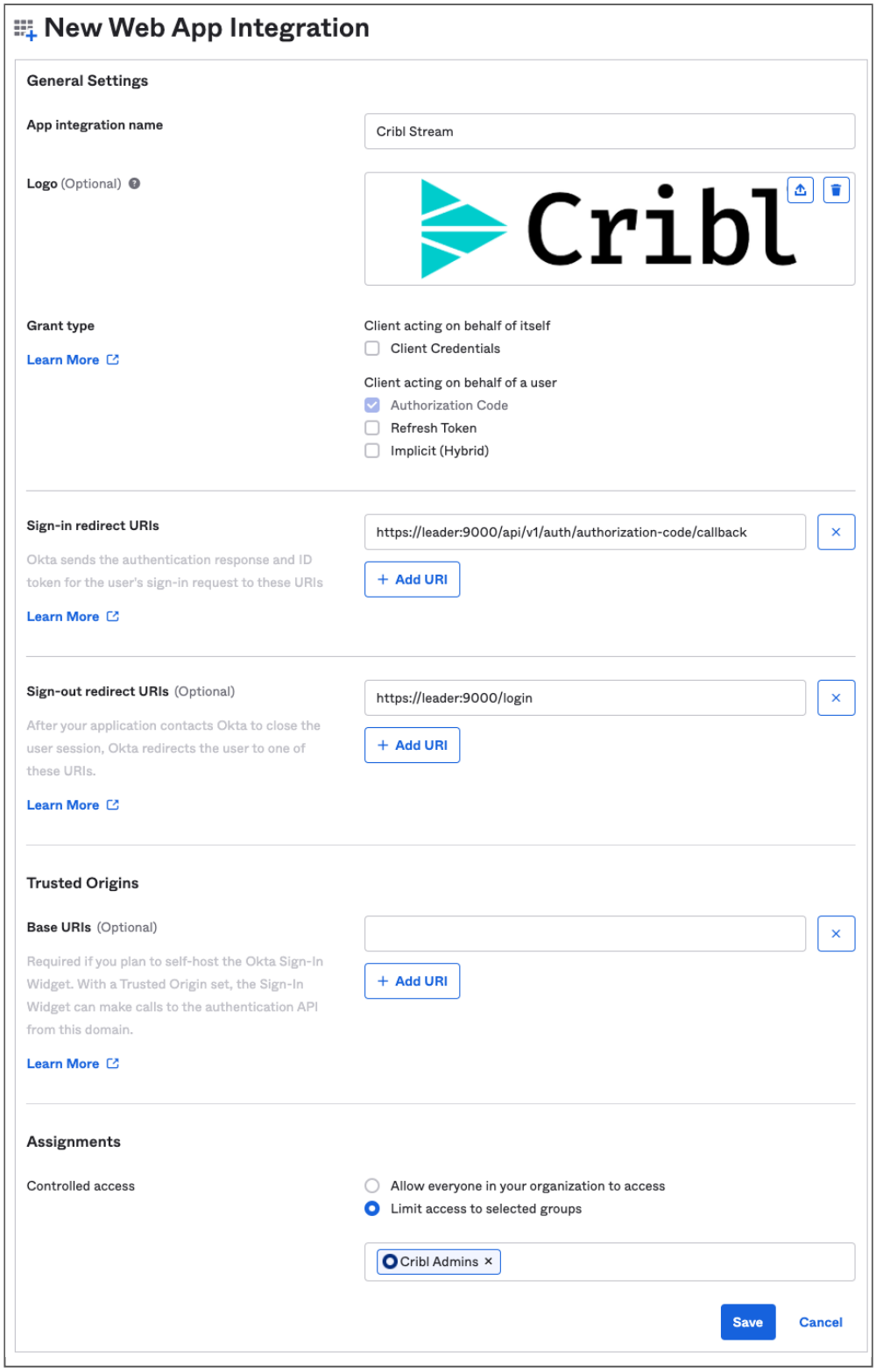
Integrate Okta with Cribl Edge

1. Log in to your Okta tenant admin console.
2. In the left nav, select **Applications** > **Applications**.
3. Click **Create App Integration**.
    - For **Sign-in method**, select `OIDC - OpenID Connect`.
    - For **Application type**, select `Web Application`.
4. Click **Next** to open the **New Web App Integration** page.
    - In the **App name** field, enter Cribl Edge.
    - (Optional) In the **Logo** field, upload the Cribl logo. You can use a logo from the Cribl Media Kit.
    - (Optional) If you wish to keep your OIDC app hidden, check the **App visibility** check box.
5. In the **Sign-in redirect URIs** field, replace the default with your Leader base URL, and with `/api/v1/auth/authorization-code/callback` as the path. This is the Cribl Edge callback API endpoint.
6. (Optional) In the **Sign-out redirect URIs** field, append `/login` to the pre-filled path.
7. In the **Assignments** > **Controlled access** area:
    - If all your Okta users need access to Cribl Edge, select **Allow everyone in your organization to access**.

- To permit specific Okta groups to access Cribl Edge, select **Limit access to selected groups**. Then, in the field below, add the groups you want to include. After you finish creating the app, if you need to add or remove groups, do that in the **Applications** > **Assignments** tab.

8. Click **Save**.

Okta should show an `Application Created Successfully` message.



Completing the new app integration in Okta

Copy Your Okta App's Client ID and Client Secret

In the **Client Credentials** panel, copy both the **Client ID** and **Client Secret**, and temporarily store them locally. You will need them in the next step, when you configure Cribl Edge.

Begin Configuring OIDC Auth in Cribl Edge

The only OAuth 2.0 flow that Cribl Edge supports is the Authorization Code Grant flow.

In version 3.0 and higher, Cribl Edge's former "master" application components are renamed "leader." Above, while some legacy terminology remains within URLs, this document will reflect that.

In Cribl Edge, select **Settings** > **Access Management** > **Authentication**.

1. Choose `OpenID Connect` from the **Type** dropdown.
2. Choose `Okta` from the **Provider** dropdown.
3. In the **Audience** field, enter your Cribl Edge UI base URL. Do **not** append a trailing slash.
4. In the **Client ID** and **Client secret** fields, enter the respective values that you copied from the Okta UI in the previous step.
5. If your Cribl Edge is in Enterprise Distributed mode:
   In the **Scope** field, add the scope `groups` to the default space-separated list of scopes, which is: `openid profile email`.
6. Obtain the authentication, token, user info, and logout URLs for your Okta app, by sending a request to the OpenID Connect Discovery endpoint.
   - This endpoint has the URL:
     `https://<tenant>.okta.com/.well-known/openid-configuration`

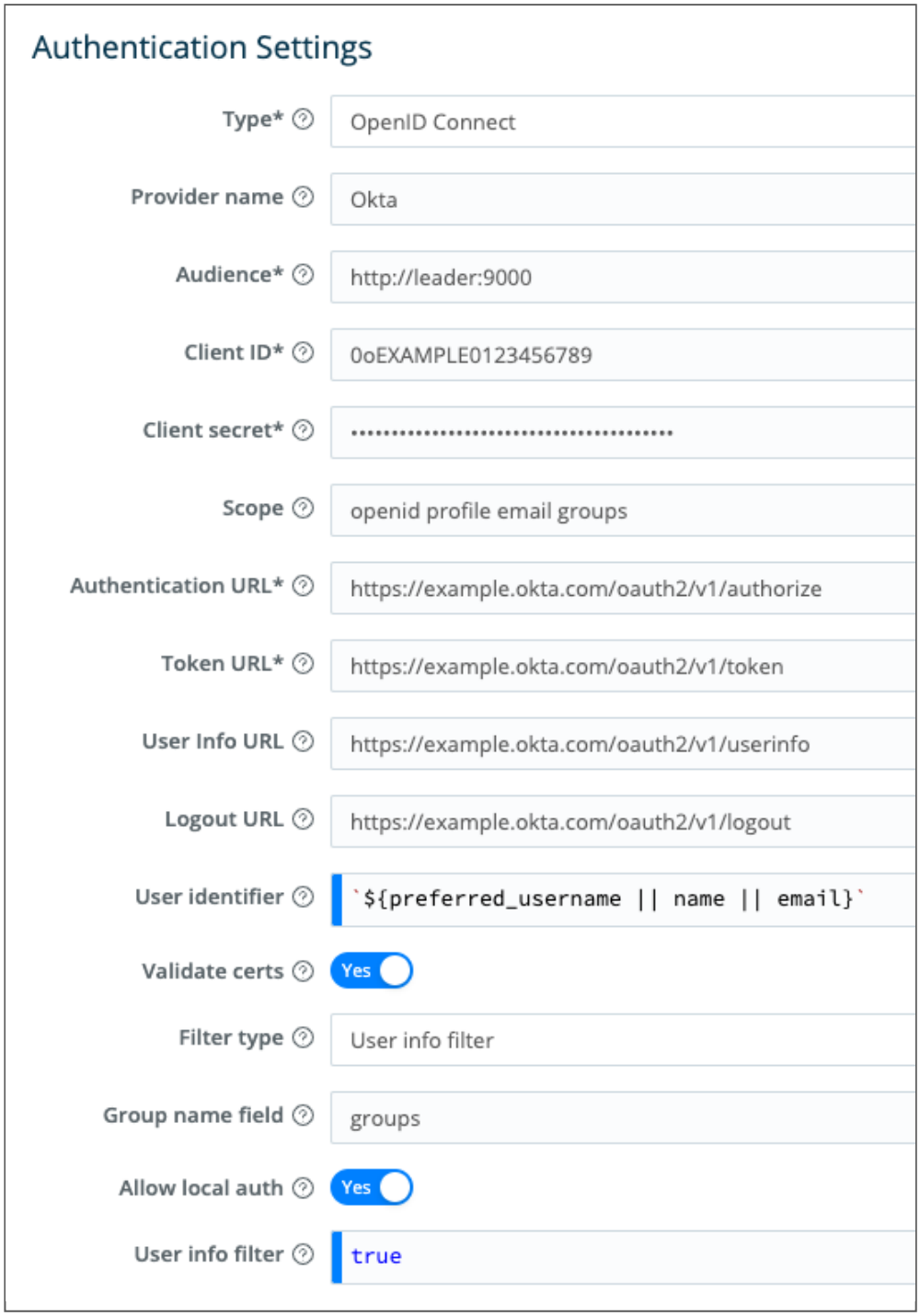
   ...where `<tenant>` is your Okta tenant name. For example:
   `https://dev-12345678.okta.com/.well-known/openid-configuration`
   - You can view the discovery document in your web browser, or use jq to extract the needed values, as in the following example:

```
curl -s https://<tenant>.okta.com/.well-known/openid-configuration | jq '. |
{"auth": (.authorization_endpoint), "token":(.token_endpoint), "userinfo":
(.userinfo_endpoint), "logout": (.end_session_endpoint)}'
```

   - Sample response:

```
{ "auth": "https://dev-416897.oktapreview.com/oauth2/v1/authorize", "token":
"https://dev-416897.oktapreview.com/oauth2/v1/token", "userinfo": "https://dev-
416897.oktapreview.com/oauth2/v1/userinfo", "logout": "https://dev-
416897.oktapreview.com/oauth2/v1/logout" }
```

7. Populate the **Authentication URL Token URL** fields with the respective `auth` and `token` URLs.
8. If you configured Okta to use groups, populate the **User info URL** field with the `userinfo` URL. This is necessary because Okta does not send group information in the `id_token` passed to Cribl Edge.
9. If you want **Account** > **Log out** in Cribl Edge to log the user out **globally**, populate the **Logout URL** field with the `logout` URL. This means that when a user clicks the **Accounts** > **Log out** link in Cribl Edge, they are logged out of **both** Cribl Edge and Okta.



Authentication settings in Cribl Edge

Configure Response to Okta `/userinfo` Endpoint

An Okta tenant's user groups can be mastered either inside Okta, outside Okta, or both.

When the `/userinfo` endpoint is queried, Okta returns the appropriate groups membership of the user back to Cribl Edge:

- For groups mastered inside Okta only, the app should pass a `Filter` type groups claim to Cribl Edge.
- For groups mastered outside Okta (such as Active Directory), or both inside and outside, the app should pass an `Expression` type groups claim back to Cribl Edge.

See the Okta documentation on dynamic allow lists and using Okta together with Active Directory.

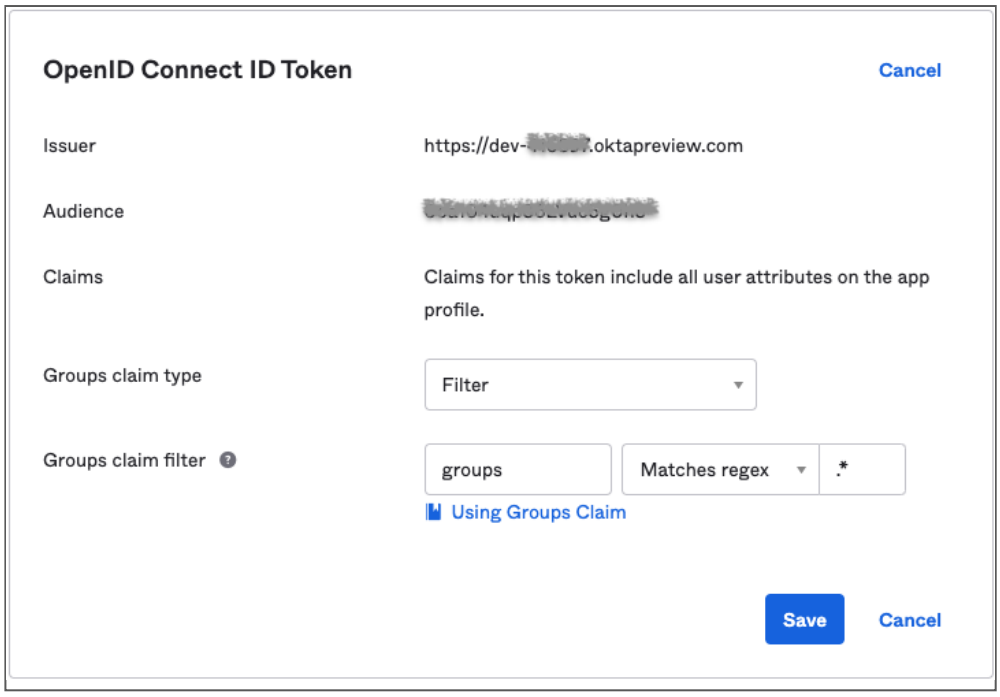
In Okta, you should still be in the panel for the app you created. If not, you can get there by opening **Applications** > **Applications** and selecting the app.

- For groups mastered **inside** Okta only, complete this procedure.
- For groups mastered **outside** Okta, or both inside and outside, complete this procedure.

Configure Groups Inside of Okta

Open the **Sign On** tab. Then, in the **OpenID Connect ID Token** panel:

1. Click **Edit** to change the value of **Groups claim filter** to `groups` and show filter options.
2. Leave **Groups claim type** set to `Filter`.
3. Choose **Matches regex** from the dropdown, and enter `.*` as the regex.
4. Click **Save**.



Role mapping, beginning

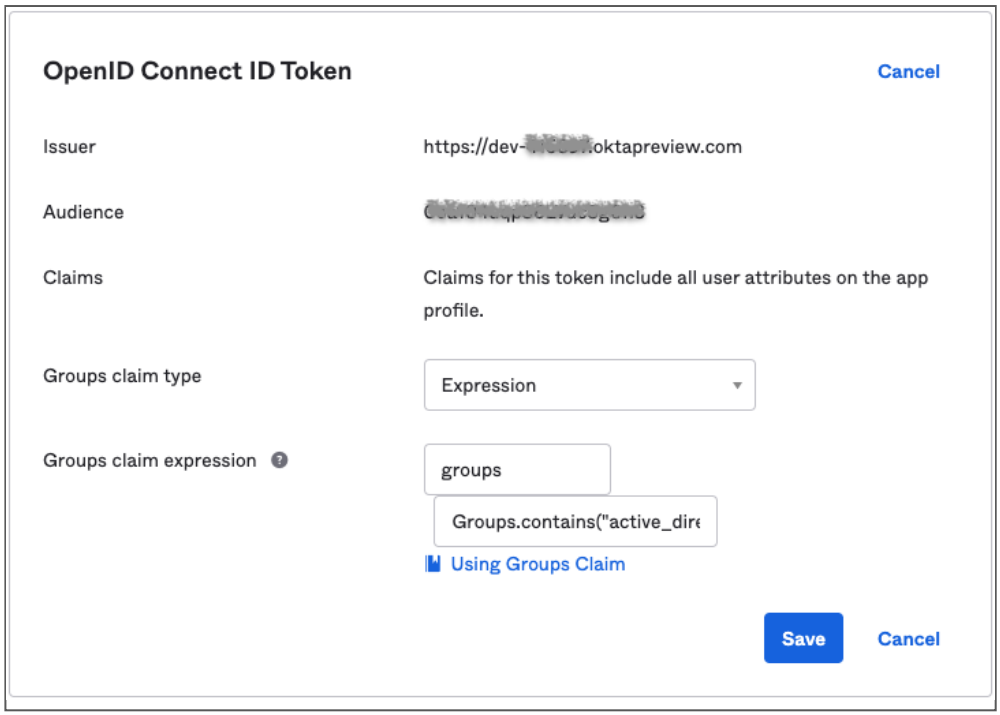
Configure Groups Outside of Okta

Open the **Sign On** tab, if necessary. Then, in the **OpenID Connect ID Token** panel:

1. Click **Edit** to change the value of **Groups claim filter** to `groups` and show filter options.
2. Set **Groups claim type** set to `Expression`.
3. In the **Groups claim expression**, enter an expression field that matches the groups you want passed to Cribl Edge. See the Okta documentation for more details.
   For example, to match on Active Directory groups that contain the string `okta`, use the following expression:

   ```
   Groups.contains("active_directory", "cribl", 10)
   ```

4. Click **Save**.

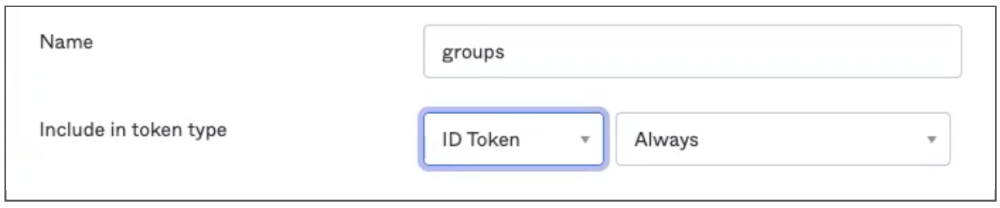


Role mapping, continued

Configure ID Token to Include Groups Claim

Okta can recognize your groups **only** if your ID token includes your groups claim, as you'll configure here.

1. In Okta, open the **Security** > **API** page.
2. In the **Authorization Servers** tab, click the edit (pencil) button for the desired Authorization Server.
3. In the resulting page, click the **Claims** tab.
4. If your groups claim already exists, click the edit (pencil) button. Otherwise, click **Add Claim**.
5. In the **Include in token type** drop-downs, choose `ID Token` and `Always`, respectively.



Including the groups claim in the token ID

6. Configure the remaining settings in the way that suits your groups claim.
7. Click **Save** (or **Create** if you're adding the claim for the first time).

Finish Configuring OIDC Auth in Cribl Edge

This section documents the entire Cribl Edge Authentication UI. If you have been working through the procedures from earlier in this page, you will have completed some of the following steps already.

Navigate to **Settings** > [**Global Settings** >] **Access Management** > **Authentication** > **Type** and select **OpenID Connect**. This exposes the following controls.

- **Provider name**: The name of the identity provider service. You can select **Google** or **Okta**, both supported natively. Manual entries are also allowed.
- **Audience (SP entity ID)**: The base URL, for example: `https://leader.yourDomain.com:9000` for a distributed environment. For the IDP, this serves as a unique identifier for the SP (that is, your Cribl Edge instance).

  For distributed environments with a second Leader configured, modify the **Audience** field to point to the load balancer instead of the Leader Node.

- **Client ID**: The `client_id` from provider configuration.
- **Client secret**: The `client_secret` from provider configuration.
- **Scope**: Space-separated list of authentication scopes. The default list is: `openid profile email`. If you populate the **User info URL** field, you must add `groups` to this list.
- **Authentication URL**: The full path to the provider's authentication endpoint. Be sure to configure the callback URL at the provider as `<masterServerFQDN>:9000/api/v1/auth/authorization-code/callback`, for example: `https://leader.yourDomain.com:9000/api/v1/auth/authorization-code/callback`.
- **Token URL**: The full path to the provider's access token URL.
- **User info URL**: The full path to the provider's user info URL. Optional; if not provided, Cribl Edge will attempt to gather user info from the ID token returned from the **Token URL**.
- **Logout URL**: The full path to the provider's logout URL. Leave blank if the provider does not support logout or token revocation.
- **User identifier**: JavaScript expression used to derive `userId` from the `id_token` returned by the OpenID provider.
- **Validate certs**: Whether to validate certificates. Defaults to `Yes`. Toggle to `No` to allow insecure self-signed certificates.
- **Filter type**: Select either **Email allowlist** or **User info filter**. This selection displays one of the following fields:
  - **Email allowlist**: Wildcard list of emails/email patterns that are allowed access.

    - **User info filter**: JavaScript expression to filter against user profile attributes. For example: `name.startsWith("someUser") && email.endsWith("domain.com")`

- **Group name field**: Field in the **User info URL** response (if configured); otherwise, `id_token` that contains the user groups. Defaults to `groups`.
- **Allow local auth**: Toggle to `Yes` to also users to log in using Cribl Edge's local authentication. This enables an extra button called `Log in with local user` on the Cribl Edge login page. (This option ensures fallback access for local users if SSO/OpenID authentication fails.)

To prevent lockout, Cribl strongly recommends enabling **Allow local auth** until you're certain that external auth is working as intended. If you do get locked out, see Manual Password Replacement.

- **Email allowlist**: Wildcard list of emails/email patterns that are allowed access.

Note the following details when filling in the form – for example, when using Okta:

- `<Issuer URI>` is the account at the identity provider.
- `Audience` is the URL of the host that will be connecting to the Issuer (for example, `https://leader.yourDomain.com:9000` for a distributed environment). The issuer (Okta, in this example) will redirect back to this site upon authentication success or failure.
- `User info URL` is required, because Okta doesn't encode groups in `id_token`. Azure AD and Google also rely on this field.

Role Mapping Settings

This section is displayed only on **distributed** deployments (Edge, Stream) with an Enterprise License. For details on mapping your external identity provider's configured groups to corresponding Cribl Edge user access Roles, see External Groups and Roles. The controls here are:

- **Default role**: Default Cribl Edge Role to assign to all groups not explicitly mapped to a Role.
- **Mapping**: Add a mapping for each external user group that you want to map to one or more Cribl roles. Then enter the group and select the role(s).

See also the explanation of role mapping above.

Role Mapping Example

Role mapping UIs will differ from one IDP to another. For this example, we'll look at Okta's.

You can assign a Cribl Edge Role to each Okta group name, and you can specify a `default` Role for users who are not in any groups.

1. In Cribl Edge, select **Settings** > **Access Management** > **Authentication**.
2. Scroll down to the **ROLE MAPPING** section.
   Cribl recommends that you set the `default` Role to `user`, meaning that this Role will be assigned to users who are not in any groups.
3. Add mappings as needed.
   The Okta group names in the left column are case-sensitive, and must match the values returned by Okta (those you saw earlier when configuring Okta and OIDC).

Role mapping, concluded

Verify that SSO with Okta Is Working

1. Log out of Cribl Edge, and verify that Okta is now an option on the login page.
2. Click **Log in with Okta**.
3. You should be redirected to Okta to authenticate yourself.
4. The OpenID connect flow should complete the authentication process.

Getting Temporary Access Credentials for AWS S3 Buckets

You can use your SSO/OIDC IDP to issue temporary access credentials so your on-prem Edge Node can access AWS S3 buckets.

Set the `AWS_WEB_IDENTITY_TOKEN_FILE` environment variable. This variable defines a path to a file that contains the OAuth/OIDC provided by the SSO IDP. You'll also need to define `AWS_ROLE_ARN` and `AWS_ROLE_SESSION_NAME`.

You can use curl/Postman to make the required API calls.

Troubleshooting Resources

Cribl University's Troubleshooting Criblet on SSO Integration for Stream On-Prem - Okta walks you through this whole configuration flow. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Troubleshooting Criblets and Advanced Troubleshooting short courses.

8.7. SSO With Okta And SAML

Cribl Edge supports SSO/SAML user authentication (login/password) and authorization (user's group membership, which you can map to Cribl Roles). Using SAML will change the default `Log in` button on the login page to a button labeled `Log in with SAML 2.0` which redirects to the configured Identity Provider (IDP).

If you are a Cribl Edge admin and want to offer single sign-on (SSO) to your users, you can choose SAML 2.0 as the authentication type, then configure an IDP to use SAML within its single sign-on flow. Once configuration is complete (several steps later), the Cribl Edge login page will send users to the IDP's login UI. Besides the IDP, some settings will refer to the Service Provider (SP), which in this context is your Cribl Edge instance.

The IDP can be Okta or Google, among others. Configuring SSO requires going back and forth between Cribl Edge and the IDP's UI. In this page, we walk through the process for configuring SSO with SAML, using Okta as the IDP.

The last part of the walkthrough is a complete description of the Cribl Edge **Authentication** settings (with SAML 2.0 selected as the authentication method). Skip directly to this section if you just need a UI reference, or if your IDP is not Okta. For non-Okta deployments, please join us on Cribl's Community Slack at https://cribl-community.slack.com/ and share your questions.

Make sure you don't get locked out of Cribl Edge! Enable the **Allow login as Local User** toggle until you're certain that external auth is working as intended. If you do get locked out, refer to Manual Password Replacement for the remedy.

SSO with SAML is supported only in Cribl Edge versions 4.1.0 and later.

Like other external auth methods, SAML requires either an Enterprise or a Standard license. It is not supported with a Free license.

Plan Your Mapping of Okta Groups to Cribl Edge Roles

In Okta, admins organize their users in groups. In Cribl Edge, there are no user groups, but there are Roles. Your task includes **mapping** Okta groups to Cribl Edge Roles.

- Mapping groups to Roles is possible only for Cribl Edge deployments that are in Distributed mode, with an Enterprise license applied.

- If you are running Cribl Edge in Single-instance mode, you cannot map Okta groups to Cribl Edge Roles, although you can still set up SSO with Okta.

As you think through how best to map your Okta groups to Cribl Edge Roles, keep these principles in mind:

- An Okta group can map to more than one Cribl Edge Role.
- A Cribl Edge Role can map to more than one Okta group.
- If a user has multiple Roles, Cribl Edge applies the union of the most permissive levels of access.
- Cribl Edge automatically assigns the `default` Role to any user who has no mapped Roles.

The example below illustrates how multiple mappings work: The groups in Mapping **b** and **c** each map to multiple Roles, while both the `reader_all` and `editor_cloud` Roles map to multiple groups.

Mapping	Okta Group	Cribl Edge Role(s)
a	Cribl Admins	`admin`
b	Cloud Admins	`reader_all`, `editor_cloud`
c	Security Team	`reader_all`, `editor_cloud`, `editor_firewall`

Cribl Edge Roles and role mapping are supported **only** with an Enterprise license. With a Standard license, all your external users will be imported to Cribl Edge in the `admin` role.

Begin Configuring SAML Auth in Cribl Edge

Navigate to **Settings** > [**Global Settings** >] **Access Management** > **Authentication** > **Type** and select **SAML 2.0**.

In the **Audience (SP entity ID)** field, enter the base URL of your Cribl Edge instance, for example, `https://yourDomain.com:9000`. Do **not** append a trailing slash.

For distributed environments with a second Leader configured, modify the **Audience** field to point to the load balancer instead of the Leader Node.

This will populate three more fields:

- **Sign-on callback URL**
- **Logout callback URL**
- **Metadata URL**

Note the values of all four fields for use in the next section.

If you want the requests that Cribl Edge sends to the IDP to be signed, select or create a **Request certificate**.

- Note the public key for the certificate that you select or create, for use in the next section.

Integrate Okta with Cribl Edge

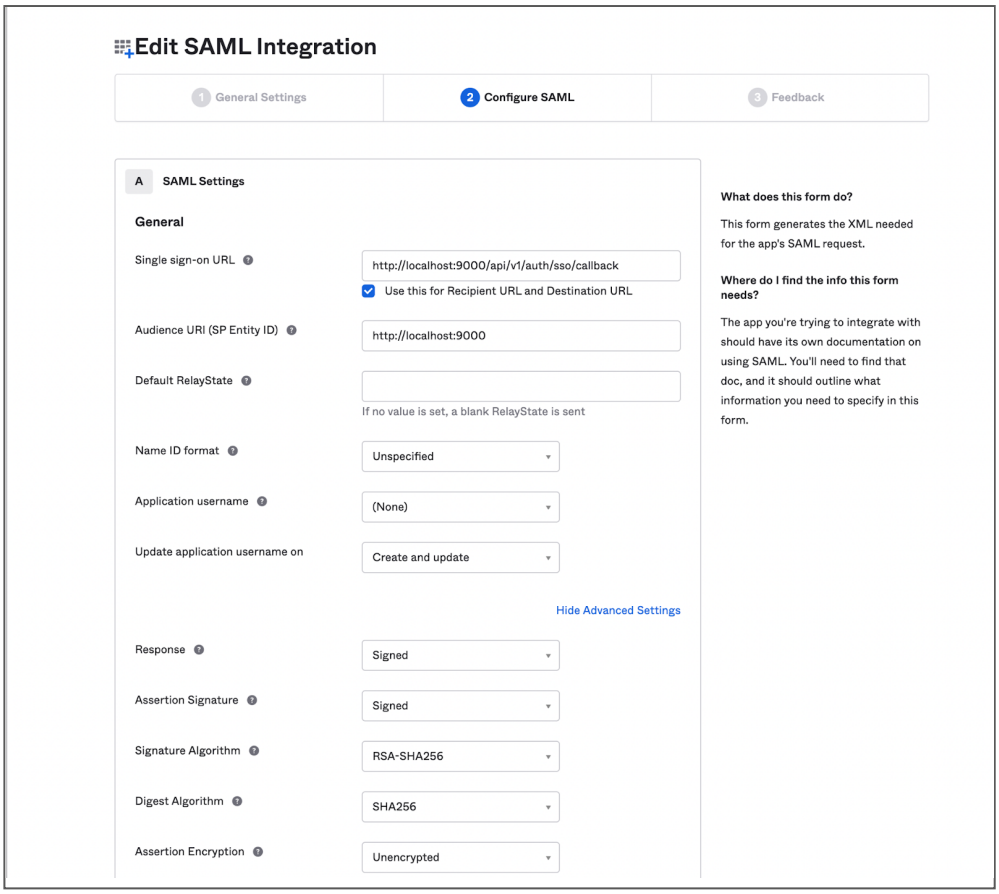
In this section, we'll create an Okta app that uses SAML and integrates with Cribl Edge.

Configure General Settings (Part 1)

1. Log in to your Okta tenant admin console.
2. In the left nav, select **Applications** > **Applications** to open the Part **1**, **General Settings** page.
3. Click **Create App Integration**.
    - For **Sign-in method**, select `SAML 2.0`.
4. Click **Next** to open the **Create SAML Integration** page.
    - In the **App name** field, enter Cribl Edge.
    - (Optional) In the **Logo** field, upload the Cribl logo. You can use a logo from the Cribl Media Kit.
    - (Optional) If you wish to keep your SAML app hidden, check the **App visibility** check box.

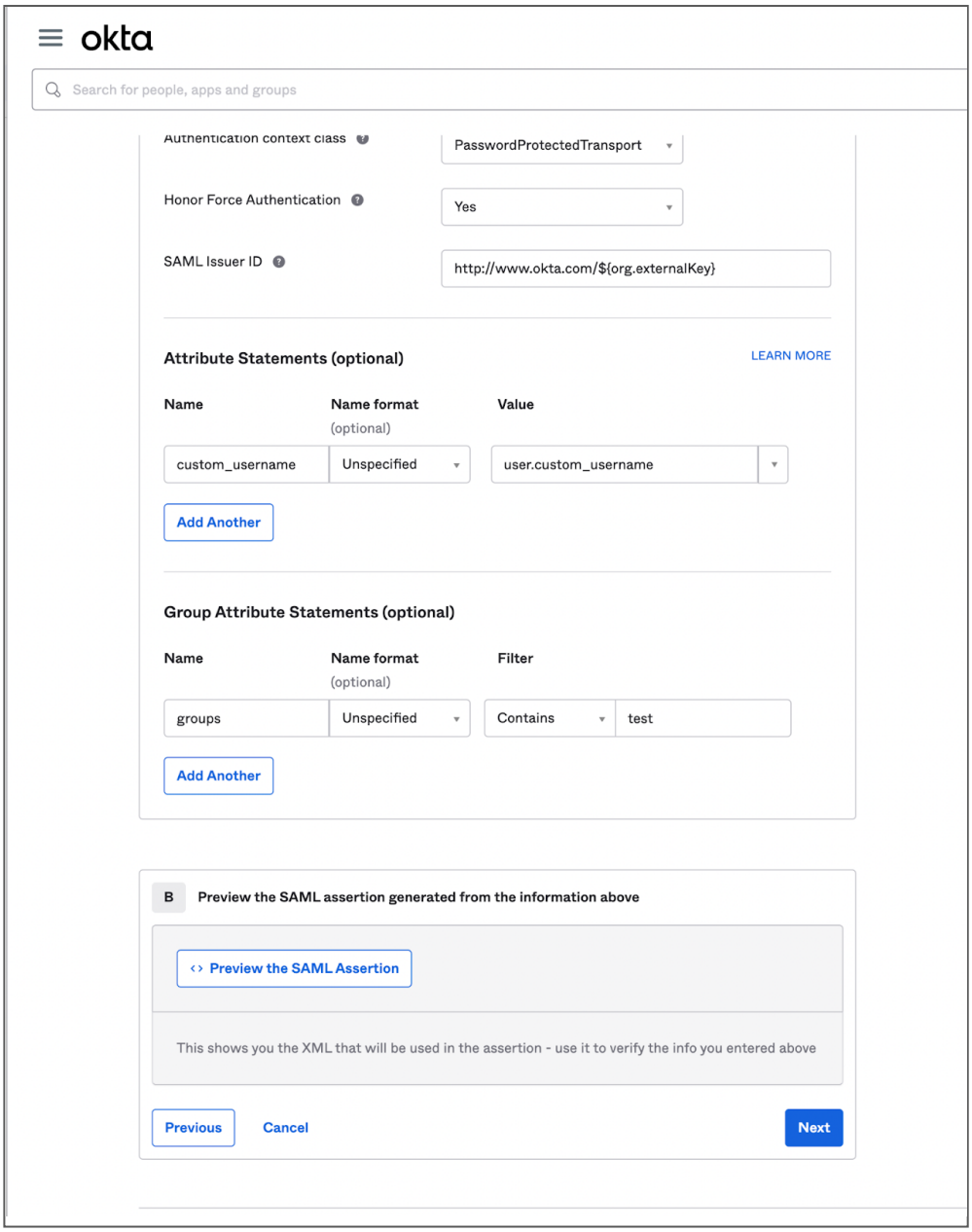
Configure SAML (Part 2)

1. Click **Next** to open the **Configure SAML** > part **A**, **SAML Settings** page. Configure these settings as follows:
    - **Single sign-on URL**: Enter the Cribl Edge **Sign-on callback URL**.
    - **Audience URI (SP Entity ID)**: Enter the Cribl Edge **Audience (SP entity ID)**.
    - **Default RelayState**: Leave blank.
    - **Name ID format**: Leave as `Unspecified` (the default).
    - **Application username**: Specify a username; can be a plain username, an email, or a custom username. In the SAML assertion's `subject` statement, this is the value for `NameID`. By default, Cribl Edge will use this value as the username in Stream. Alternatively, you can set a custom attribute statement in Okta (as described in Step 2 below), then set the **Username field** in Cribl Edge to use that instead.
    - **Update application username on**: Leave as is (`Create and update`).
    - (Optional) Click **Show Advanced Settings** if you want to configure Single Logout, SAML response encryption, etc., as described below.



Configuring SAML (Part 2)

2. (Optional) Define custom **Attribute Statements** that Okta will insert in the SAML Assertions shared with Cribl Edge. The only use case that Cribl Edge supports for this setting is creating a custom **Application username**, as described in the previous step.
3. (Optional) Configure **Group Attribute Statements**. Similar to the previous step, except that here, Cribl Edge supports creating a custom attribute whose value is one or more Okta groups that will populate Cribl Edge's **Group name field**.
4. In Part **B**, if you want to see your SAML assertion in XML form, click **Preview the SAML assertion generated from the information above**.



Configuring SAML (Part 2) continued

Copy Your App's Metadata to Cribl Edge (Part 3)

1. Click **Next** to open the **Feedback** page, whose fields are self-explanatory, and click **Finish**. Okta takes you to the **Sign On** tab for your newly-created application.
2. Under **SAML Setup** in the right margin, click **View SAML setup instructions** to open the **How to Configure SAML 2.0 ...** page.
3. Copy each value below to its corresponding field in Cribl Edge. Values that you did not configure above will not appear on the page.

SAML App Value	Cribl Edge Field
Identity Provider Single Sign-On URL	**Single sign-on (SSO) URL**

SAML App Value	Cribl Edge Field
Identity Provider Single Logout URL	Single logout (SLO) URL
Identity Provider Issuer	Issuer (IDP entity ID)
X.509 Certificate	Response validation certificate

Ignore the **Provide the following IDP metadata to your SP provider** text box. Cribl Edge does not support ingesting metadata in this form.

Advanced Settings

The **Configure SAML** > part **A**, **SAML Settings** page offers these **Advanced Settings**:

Response: This is the SAML response object; it contains an `Assertion` sub-field. Leave this setting as `Signed` (the default), because this is the only alternative that Cribl Edge supports.

Assertion Signature: This applies to the `Assertion` sub-field within the SAML response object. Defaults to `Signed`, but that has no practical effect because Cribl Edge assumes that the whole SAML response object is signed anyway.

Signature Algorithm Select your preferred algorithm for signing the response that the IDP sends to Cribl Edge.

Digest Algorithm: Select your preferred hashing algorithm for the response that the IDP sends to Cribl Edge.

Assertion Encryption: Choose `Encrypted` if you want the IDP to encrypt the response that it sends to Cribl Edge. Setting this to `Encrypted` displays the following three additional options:

- **Encryption Algorithm**: Select your preferred algorithm for encrypting the SAML response.
- **Key Transport Algorithm**: Select your preferred algorithm for encrypting the encryption key itself.
- **Encryption Certificate**: Upload the public key of the Cribl Edge **Response decryption certificate** you selected or created in Cribl Edge.

Signature Certificate: Upload the public key of the Cribl Edge **Request certificate** you selected or created in Cribl Edge.

Enable Single Logout: Check the checkbox if you want Cribl Edge to send a logout request to Okta upon logout of the Cribl Edge user. Doing this displays the following two additional options that use values from earlier in these instructions:

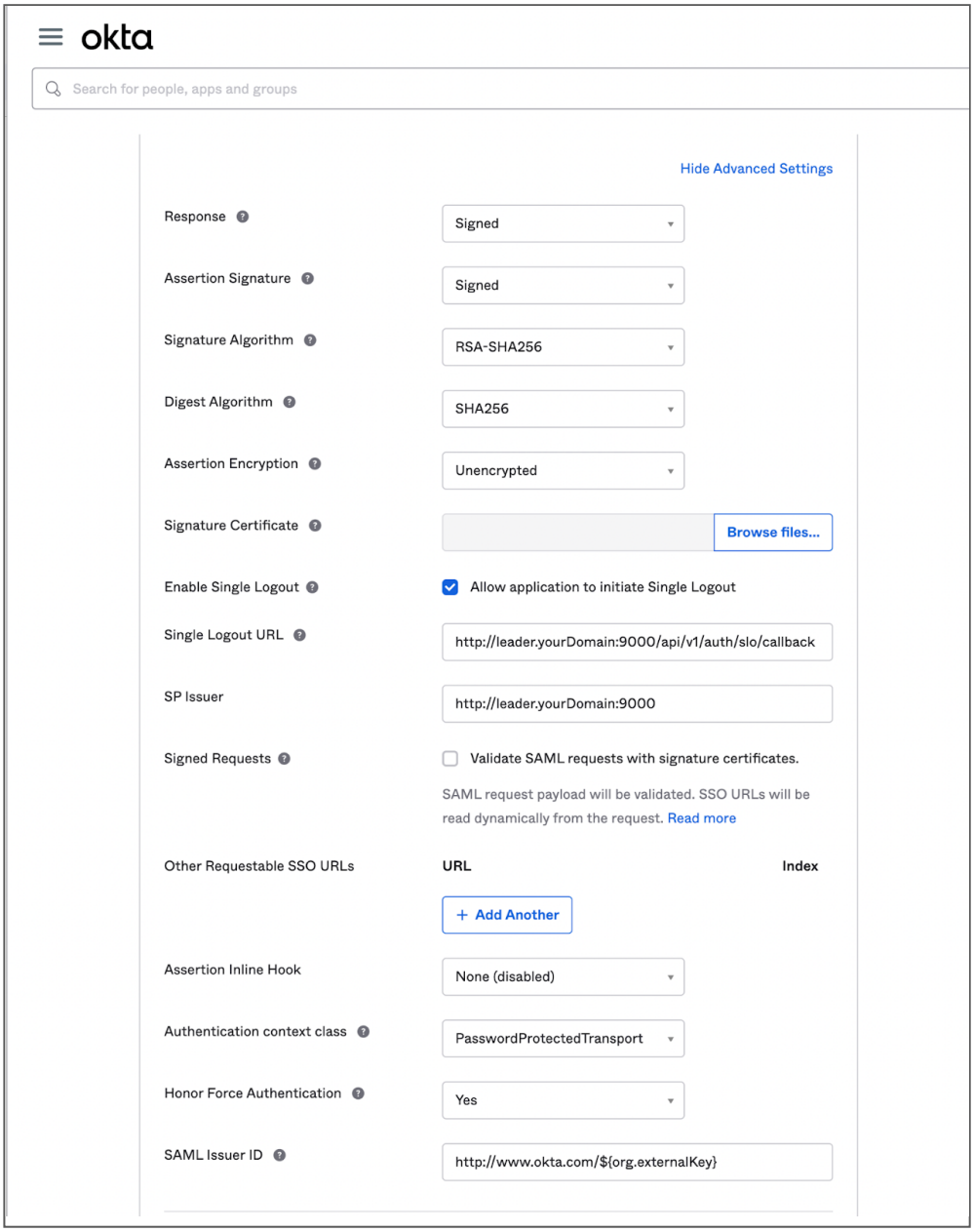
- **Single Logout URL**: Enter the value that Cribl Edge provided for **Logout callback URL**.

- **SP Issuer**: Enter the URL that you entered for the **Audience** setting in the Cribl Edge UI.

Signed Requests: Check the checkbox if you want Okta to validate the signatures of SAML requests that come from Cribl Edge.

Other Requestable SSO URLs, **Assertion Inline Hook**, **Authentication context class**, and **Honor Force Authentication**: Ignore these settings (which are not supported by Cribl Edge).

SAML Issuer ID: Enables you to override Okta's default value for **Identity Provider Issuer**.



Okta Advanced Settings

Finish Configuring SAML Auth in Cribl Edge

This section documents the entire Cribl Edge Authentication UI. If you have been working through the procedures from earlier in this page, you will have completed some of the following steps already.

Navigate to **Settings** > [**Global Settings** >] **Access Management** > **Authentication**. This exposes the following controls.

Audience (SP entity ID): The base URL, for example: `https://leader.yourDomain.com:9000` for a distributed environment. For the IDP, this serves as a unique identifier for the SP (that is, your Cribl Edge instance).

For distributed environments with a second Leader configured, modify the **Audience** field to point to the load balancer instead of the Leader Node.

Once you enter a URL here, Cribl Edge will automatically populate the following three read-only fields. Copy the values for use in your IDP's UI.

- **Sign-on callback URL**: URL where the SP (Cribl Edge) will consume assertions (that is, will receive SAML authentication responses from the IDP).
- **Logout callback URL**: URL where the SP (Cribl Edge) will receive SAML logout responses from the IDP.
- **Metadata URL**: URL that exposes a description of the SP (Cribl Edge), including endpoints, certificates, and metadata. Typically, you paste this URL into an **import metadata from URL** or similar field during setup of IDPs that support such an option.

Request certificate: Certificate the SP (Cribl Edge) should use to sign requests. Create a new certificate, or choose one that you imported as described here. Typically pasted into a **Signature Certificate**, **Verification Certificate**, or similar field during IDP setup. Required by some but not all IDPs; for example, when SLO is enabled, Okta requires signed requests.

Issuer (IDP entity ID): Unique ID of the IDP. In the IDP's UI, typically called **Identity Provider Issuer**, **IDP Identifier**, or similar.

Single sign-on (SSO) URL: Endpoint to which authentication requests will be sent (that is, the IDP's single sign-on service). In the IDP's UI, typically called **Login URL**, **Single Sign-On URL**, or similar.

Single logout (SLO) URL: Endpoint to which logout requests will be sent (that is, IDP's single logout service). Setting this will enable single logout initiated from the SP (Cribl Edge).

Request binding: Type of binding for SAML requests sent to the identity provider's SSO/SLO services.

Response validation certificate: The certificate that the SP (Cribl Edge) should use to validate signed responses. Provided by the IDP; contains a public key.

If, in Okta, you have set **Assertion Encryption** to `Encrypted`, then, in Cribl Edge you must both specify a certificate in **Response decryption certificate** below, and, set **Order of encrypting and signing** to `encrypt-then-sign`.

Response decryption certificate: Certificate the SP (Cribl Edge) should use to decrypt encrypted responses. Obtain this from the IDP, where it is typically labeled **Token Encryption**, **Encryption Certificate**, or similar. Create a new certificate, or choose one that you imported as described here.

Order of encrypting and signing: Specify the order in which the IDP signs and encrypts responses, if you've configured it to do that. Requires a **Response decryption certificate** to be set.

Username field: In the SAML response, the element from which Cribl Edge should extract the user identifier.

- Defaults to `NameID`, which is an element in the `Subject` block, for example:

```
<saml:Subject>
    <saml:NameID Format="urn:oasis:names:tc:SAML:1.1:nameid-format:unspecifiec
    <saml:SubjectConfirmation Method="urn:oasis:names:tc:SAML:2.0:cm:bearer">
        <saml:SubjectConfirmationData NotOnOrAfter="2014-02-22T01:20:27.956Z"
            Recipient="http://localhost/ExampleServiceProvider/SAML/AssertionC
    </saml:SubjectConfirmation>
</saml:Subject>
```

- Alternatively, if you want to set the username via a custom Attribute when configuring the IDP, you can use the `AttributeStatement` block, for example:

```
<AttributeStatement>
    <Attribute Name="urn:oid:2.5.4.42" FriendlyName="givenName" NameFormat="ur
        <AttributeValue>Dave</AttributeValue>
    </Attribute>
</AttributeStatement>
```

- When you are using a custom Attribute as in the previous bullet point, the value for **Username field** must be what's in the `Name` field of the Attribute – typically an OID-style name (like the example above, where it's `urn:oid:2.5.4.42`). Do **not** use the `FriendlyName`.

Group name field: In the SAML response, the Attribute that specifies the user groups. Defaults to "groups". In the IDP's UI, typically labeled **Group Attribute Statements** or similar.

- When configuring the **Group Attribute Statements** or equivalent in your IDP, view the full Attribute. You should see something like this:

```
<saml2:AttributeStatement xmlns:saml2="urn:oasis:names:tc:SAML:2.0:assertion">
    <saml2:Attribute Name="groups" NameFormat="urn:oasis:names:tc:SAML:2.0:att
        <saml2:AttributeValue xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlr
            test-admin
        </saml2:AttributeValue>
        <saml2:AttributeValue xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlr
            test-group
        </saml2:AttributeValue>
    </saml2:Attribute>
</saml2:AttributeStatement>
```

- Note the value for the `Attribute Name`. In the example above, it's `groups`.
- Enter that exact value as the **Group name field** in Cribl Edge.

Allow login as Local User: Toggle to Yes to allow users to log in without SAML (as an alternative, or as a fallback if SAML login fails).

Role Mapping Settings

This section is displayed only on **distributed** deployments (Edge, Stream) with an Enterprise License. For details on mapping your external identity provider's configured groups to corresponding Cribl Edge user access Roles, see External Groups and Roles. See also the explanation of role mapping above. The controls here are:

Default role: Default Cribl Edge Role to assign to all groups not explicitly mapped to a Role.

Mapping: Add a mapping for each external user group that you want to map to one or more Cribl roles. Then enter the group and select the role(s).

See also the explanation of role mapping above.

Role Mapping Example

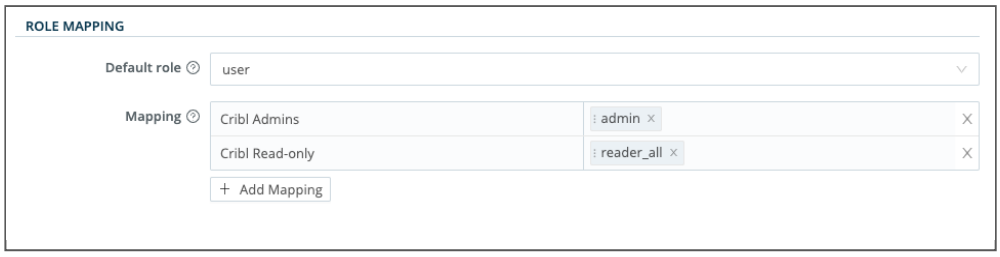
Role mapping UIs will differ from one IDP to another. For this example, we'll look at Okta's.

You can assign a Cribl Edge Role to each Okta group name, and you can specify a `default` Role for users who are not in any groups.

1. In Cribl Edge, select **Settings** > **Access Management** > **Authentication**.
2. Scroll down to the **ROLE MAPPING** section.

Cribl recommends that you set the `default` Role to `user`, meaning that this Role will be assigned to users who are not in any groups.

3. Add mappings as needed.
   The Okta group names in the left column are case-sensitive, and must match the values returned by Okta (those you saw earlier when configuring Okta and SAML).



Role mapping, concluded

Verify that SSO with SAML Is Working

1. Log out of Cribl Edge, and verify that SAML is now an option on the login page.

Logging in with SAML

2. Click **Log in with SAML 2.0**.
3. You should be redirected to your IDP to authenticate yourself.
4. The SSO connect flow should complete the authentication process.

Getting Temporary Access Credentials for AWS S3 Buckets

You can use your SSO/SAML IDP to issue temporary access credentials so your on-prem Edge Node can access AWS S3 buckets.

Call the `AssumeRoleWithSAML` API endpoint. It will return the STS access, secret, and session tokens. These can be written into the `~/.aws/` credentials file, which Cribl Edge will pick up because it uses the native AWS SDK.

You can set up multiple S3 Sources with different credentials. Cribl Edge relies on the AWS SDK for authentication support, and the SDK evaluates credentials in the following order:

1. Loaded from AWS Identity and Access Management (IAM) roles for Amazon EC2.
2. Loaded from the shared credentials file (~/.aws/credentials).
3. Loaded from environment variables.
4. Loaded from a JSON file on disk.
5. Other credential-provider classes provided by the JavaScript SDK.

To enable the use of multiple Sources, set the S3 Source **Authentication method** to **Auto**.

See the AWS SDK documentation and AWS CLI documentation for further information on setting credentials.

9. Securing

9.1. Leader's Auth Token

For new installations, Cribl Edge generates a secure, random Leader's auth token. A new token is also created when you switch from Single-instance to the Distributed mode by configuring a node to be a Leader mode.

In versions before v4.5.0, the default token was `criblmaster`. It is not changed if you upgrade your deployment to v4.5.0 or later.

If you prefer to use your own token (or you are still using the older default of `criblmaster`), change the token value to a unique secure value.

This applies whether you configure your Leader via the UI, or via a Docker Compose file.

Changing the Auth Token Value

Changing the auth token once you have Edge Nodes deployed will break communication between the nodes and the Leader.

We recommend changing the auth token *before* deploying Edge Nodes.

Create a unique, strongly secure token value. Cribl recommends creating a string that contains at least 14 characters and includes uppercase letters, lowercase letters, and numbers.

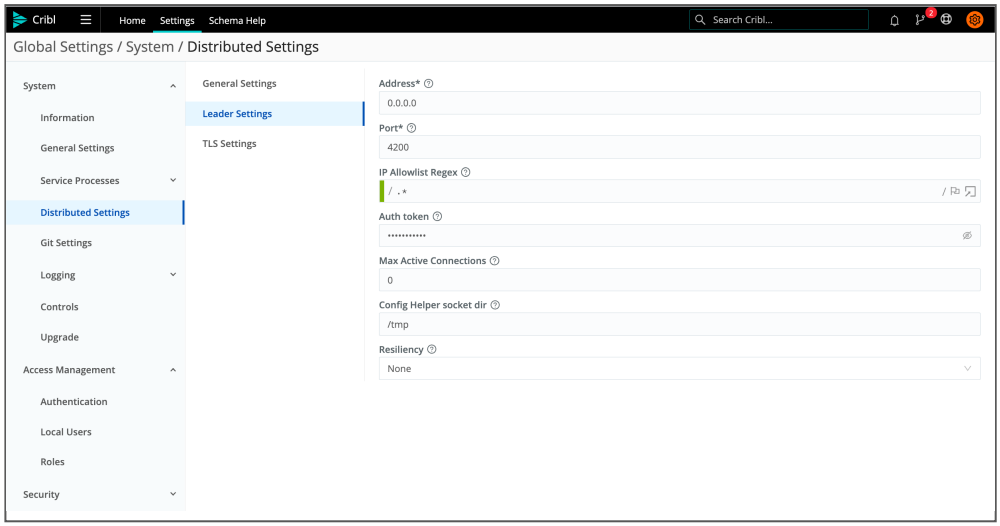
One option is to generate a token value with a shell command. For example, on Linux or Mac OS, the following command generates a 32-character value:

```
head /dev/urandom | LC_ALL=C tr -dc A-Za-z0-9 | head -c32 | cut -c 1-
```

Once you have created a secure auth token value, you can change it:

- in the **Auth token** field in the UI
- in the appropriate place in your Docker Compose file (if applicable)
- via command line: `./cribl mode-master -u <token>`

In the UI, you'll find the **Auth token** setting under **Global Settings** > **System** > **Distributed Settings** > **Leader Settings**, as shown in the screenshot below.



Distributed > Leader Settings

9.2. Securing Cribl Edge (TLS/SSL, Certs, Keys)

You can secure Cribl Edge access and traffic using various combinations of SSL (Secure Sockets Layer), TLS (Transport Layer Security), custom HTTP headers, and internal or external KMS (Key Management Service) options. An additional option for Cribl Stream is to deploy in FIPS mode.

In a single-instance deployment (Stream, Edge), wherever this page refers to a Fleet, just follow the Cribl Edge top nav's **Manage** link.

Secure Access to Edge Nodes' UI

A best practice in enterprise distributed deployments, this prevents direct browser access to Edge Nodes' UI.

Cribl recommends that you first enable the Leader's UI access to Edge Nodes (Stream, Edge). This way, admins will still be able to tunnel through from the Leader to any Edge Node's UI. This is also a prerequisite for Connecting Workers to the Leader Securely.

1. Select a Fleet.
2. Open **Fleet Settings** (top right).
3. Under **General Settings** > **API Server Settings**, click **Advanced**.
4. Toggle **Local UI Access** to `No`.
5. Click **Save**.

SSL Certificate Configuration

You can secure Cribl Edge's API and UI access by configuring SSL. Do this on the Leader, to secure Edge Nodes' inbound communications.

You can use your own certs and private keys, or you can generate a pair with OpenSSL, as shown here:

```
openssl req -nodes -new -x509 -newkey rsa:2048 -keyout myKey.pem -out myCert.pem -days 420
```

This example command will generate both a self-signed cert `myCert.pem` (certified for 420 days), and an unencrypted, 2048-bit RSA private key `myKey.pem`. (Change the filename arguments to modify these placeholder names.)

As indicated by these examples, Cribl Edge expects certificates and keys to be formatted in privacy-enhanced mail (`.pem`) format.

In the Cribl Edge UI, you can configure the cert via **Settings** > **Global Settings** > **Security** > **Certificates**. You can configure the **key** via:

- **Settings** > **Global Settings** > **Security** > **Encryption Keys** single-instance deployments (Edge, Stream), or
- **Manage** > **Groups** > `<group-name>` > **Group Settings** > **Security** > **Encryption Keys** distributed deployments (Edge, Stream).

Alternatively, you can edit the `local/cribl.yml` file's `api` section to directly set the `privKeyPath` and `certPath` attributes. For example:

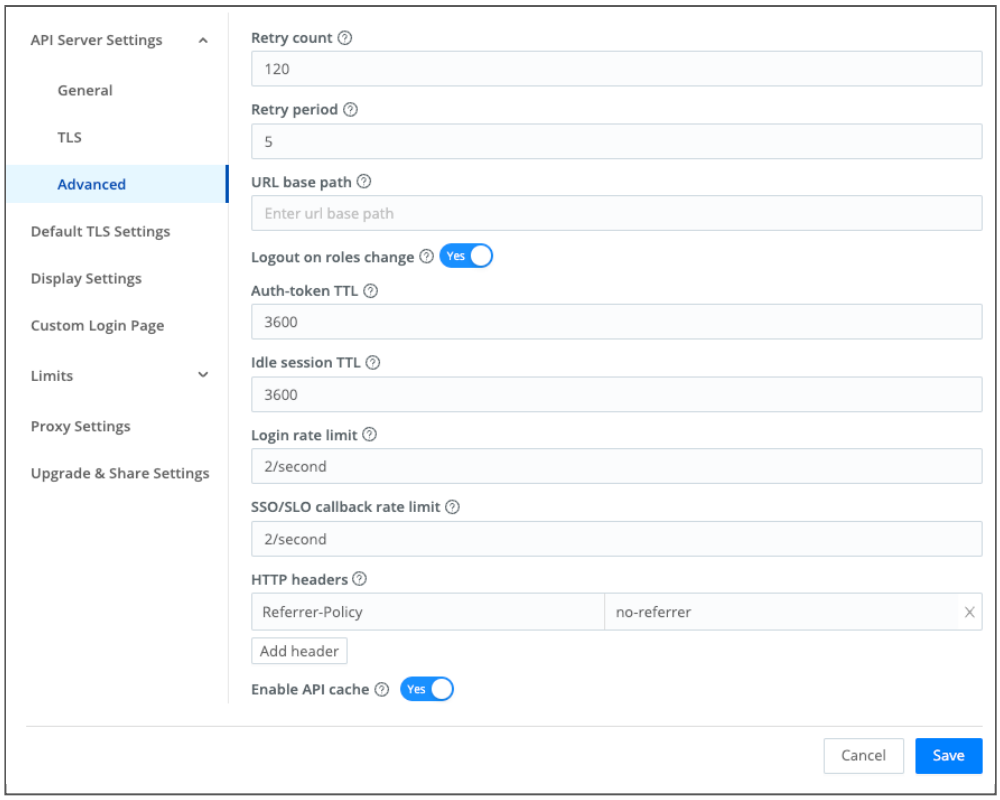
cribl.yml

```
api:
  host: 0.0.0.0
  port: 9000
  disabled : false
  ssl:
    disabled: false
    privKeyPath: /path/to/myKey.pem
    certPath: /path/to/myCert.pem
...
```

See Securing Communications for details about using this certificate and key to secure communications on, and among, your Cribl Edge Leader and Edge Nodes.

Custom HTTP Headers

You can encode custom, security-related HTTP headers, as needed. As shown in the examples below, you specify these at **Settings** > **Global Settings** > **General Settings** > **API Server Settings** > **Advanced** > **HTTP Headers**. Click **Add Header** to display extra rows for new key-value pairs.



Custom HTTP headers

TLS Settings and Traffic Types

This table shows TLS client/server pairs, and encryption defaults, per traffic type.

Traffic Type	TLS Client	TLS Server	Encryption	Cert Auth	CN* Check
UI	Browser	Cribl Edge	Default disabled	Default disabled	Default disabled
API	Worker/Edge Node	Leader	Default disabled	Default disabled	Default disabled
Worker-to-Leader	Worker/Edge Node	Leader	Default disabled	Default disabled	Default disabled
Data	Any data sender	Cribl Edge (Source)	Default disabled	Default disabled	Default disabled
Data	Cribl Edge (Destination)	Any data receiver	Default disabled	Default disabled	Default disabled
Authentication	————	————	————	————	————
Local*	Browser	Cribl Edge	Default Disabled	N/A	N/A

Traffic Type	TLS Client	TLS Server	Encryption	Cert Auth	CN* Check
LDAP*	Cribl Edge	LDAP Provider	Custom	N/A	Default Disabled
Splunk*	Cribl Edge	Splunk Search Head	Default Enabled	N/A	Default Disabled
OIDC†/Okta*	Browser and Cribl Edge	Okta	Default Enabled	N/A	Enabled (Browser)
OIDC†/Google*	Browser and Cribl Edge	Google	Default Enabled	N/A	Enabled (Browser)

** Common name*
† OpenID Connect

Default TLS Settings (Cyphers, Etc.)

You can configure advanced, system-wide TLS settings – minimum and maximum TLS versions, default cypher lists, and ECDH curve names. Select **Settings** > **Global Settings** > **System** > **General Settings** > **Default TLS Settings**.

Here, in Cribl Edge's **Default cypher list** field, you can specify between one and all of the following supported cyphers:

- `TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384`
- `TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256`
- `TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384`
- `TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256`

TLS in Cribl.Cloud

TLS encryption is pre-enabled on several Sources in Cribl.Cloud, indicated on the Cribl.Cloud portal's **Data Sources** tab. All TLS is terminated by individual Nodes.

To enable TLS settings for additional Sources, use these configuration settings:

- **Private key path:** `/opt/criblcerts/criblcloud.key`
- **CA certificate path**: `/opt/criblcerts/criblcloud.crt`
- **Minimum TLS version**: `TLSv1.2`

For steps to enable TLS mutual authentication in Cribl.Cloud, see Securing Communications.

Encryption Keys

You can create and manage keys that Cribl Edge will use for real-time encryption of fields and patterns within events. For details on applying the keys that you define here, see Encryption.

In Cribl Edge 4.1 and later, for enhanced security, Cribl encrypts TLS certificate private keys in configuration files when you add or modify them. Before upgrading to v.4.1.0 (or later), back up your existing unencrypted key files – see Safeguarding Unencrypted Private Keys for Rollback. If you need to roll back to a pre-4.1 version, see Restoring Unencrypted Private Keys below.

Accessing Keys

- In a single-instance deployment, select **Settings** > **Security** > **Encryption Keys**.
- In a distributed deployment with one Fleet, select **Settings** > **Security** > **Encryption Keys**.
- In a distributed deployment with multiple Fleets, keys are managed per Fleet. Select **Manage** > **Groups** > `<group-name>` **Group Settings** > **Security** > **Encryption Keys**.

On the resulting **Manage Encryption Keys** page, you can configure existing keys, and/or use the following options to add new keys.

Get Key Bundle

To import existing keys, click **Get Key Bundle**. You'll be prompted to supply a login and password to proceed.

Add New Key

To define a new key, click **New Key**. The resulting **New Key** modal provides the following controls:

Key ID: Cribl Edge will automatically generate this unique identifier.

Description: Optionally, enter a description summarizing this key's purpose.

Encryption algorithm: Currently, Cribl Edge supports the `aes-256-cbc` (default) and `aes-256-gcm` algorithms.

KMS for this key: Currently, the only option supported here is `local` (Cribl Edge's internal Key Management Service).

Key class: Classes are arbitrary collections of keys that you can map to different levels of access control. For details, see Encryption. This value defaults to `0`; you can assign more classes, as needed.

Expiration time: Optionally, assign the key an expiration date. Directly enter the date or select it from the date picker.

Use initialization vector: If enabled, Cribl Edge will seed encryption with a nonce to make the key more random and unique. Optional (and defaults to disabled) with the `aes-256-cbc` algorithm; automatically enabled (and cannot be disabled) with the `aes-256-gcm` algorithm.

Initialization vector size: Length of the initialization vector (IV), in bytes. This option is displayed only with the `aes-256-gcm` algorithm. Defaults to `12` bytes to optimize interoperability, but you can use the drop-down to set this anywhere between `12–16` bytes.

Restoring Unencrypted Private Keys

If you need to roll back from Cribl Edge 4.1 or later to an earlier version, follow this procedure to restore the unencrypted private key files that you earlier backed up.

1. Stop Cribl Edge 4.1 (or later) on hosts – the Leader and all Edge Nodes.
2. Untar the older Cribl Edge version (e.g., 4.0.4) onto the Leader and all Edge Nodes.
3. Manually edit each Cribl Edge config file that contains an encrypted private key. Replace each key with its prior unencrypted version.
4. Start up Cribl Edge, commit and deploy, and resume normal operations.

Secrets

With Cribl Edge's secrets store, you can centrally manage secrets that Cribl Edge instances use to authenticate on integrated services. Use this UI section to create and update authorization tokens, username/password combinations, and API-key/secret-key combinations for reuse across the application.

Accessing Secrets

- In a single-instance deployment, select **Settings** > **Security** > **Secrets**.
- In a distributed deployment with one Fleet, select **Configure** > **Settings** > **Security** > **Secrets**.
- In a distributed deployment with multiple Fleets, secrets are managed on each Fleet. Select **Groups** > `<group-name>` **Settings** > **Security** > **Secrets**.

On the resulting **Manage Secrets** page, you can configure existing secrets, and/or click **New Secret** to define new secrets.

Add New Secret

The **New Secret** modal provides the following controls:

Secret name: Enter an arbitrary, unique name for this secret.

Secret type: See below for this second field's options, some of which expose additional controls.

Description: Optionally, enter a description summarizing this secret's purpose.

Tags: Optionally, enter one or multiple tags related to this secret.

Secret Type

This drop-down offers the following types:

Text: This default type exposes a **Value** field where you directly enter the secret.

API key and secret key: Exposes **API key** and **Secret key** fields, used to retrieve the secret from a secure endpoint. This is the only secret type supported on Cribl Edge's AWS-based Sources, Collectors, and Destinations, and on our Google Cloud Storage Destination.

Username with password: Exposes **Username** and **Password** fields, which you fill to retrieve the secret using Basic Authentication.

CA Certificates and Environment Variables

Where Cribl Edge Sources and Destinations support TLS, each Source's or Destination's configuration provides a **CA Certificate Path** field where you can point to corresponding Certificate Authority (CA) `.pem` file(s). However, you can also use environment variables to manage CAs globally. Here are some common scenarios:

1. **How do I add a set of trusted root CAs to the list of trusted CAs that Cribl Edge trusts?**
   Set this environment variable in each Edge Node's environment (e.g., in its systemd unit file): `NODE_EXTRA_CA_CERTS=/path/to/file_with_certs.pem`. For details, see the nodejs docs.
2. **How do I make Cribl Edge trust all TLS certificates presented by any server it connects to?**
   Set this environment variable: `NODE_TLS_REJECT_UNAUTHORIZED=0` – for details, see the nodejs docs.

9.2.1. KMS Configuration

Cribl Edge's Key Management Service (KMS) maintains the keys that Cribl Edge uses to encrypt secrets on Fleets and Edge Nodes. The internal KMS is always available, but integrating an external KMS provider requires an Enterprise or Standard license.

To configure KMS for Cribl Edge in Cribl.Cloud, see the Launch Guide.

In an on-prem single-instance deployment, you configure the KMS at **Settings** > **Global Settings** > **Security** > **KMS**. In a distributed deployment, you configure the Leader's KMS at the same global location, while additional KMS configs for each Fleet are available at the Fleet's **Group Settings** > **Security** > **KMS** page.

The resulting **KMS Provider** drop-down currently provides these options:

- Stream Internal: The **only** option available without an Enterprise license/plan. With this option, the secrets themselves are configured and maintained in Cribl Edge Settings' parallel Secrets section.
- HashiCorp Vault
- AWS KMS

External KMS Providers and Fleets

To integrate an external KMS provider into an on-prem distributed deployment, Cribl Edge's Leader Node must have internet access.

When you initially install a license in distributed mode, a known bug prevents immediate use of KMS features within Fleets. Here is the workaround:

1. Open **Settings** > **Global Settings** > **Worker Processes**.
2. In the list of processes, locate any process with a Role of `CONFIG_HELPER`.
3. Click that process' **Restart** button.

Upon restarting, KMS will be available for use in the corresponding Fleet.

Internal KMS

The **KMS provider** field defaults to `Stream Internal`. With this option, no further configuration here is required (or possible). See Secrets to configure individual secrets.

HashiCorp Vault

Setting the **KMS provider** drop-down to `HashiCorp Vault` exposes the following configuration options:

KMS Settings

Vault URL: Enter the Vault server's URL (e.g., `http://localhost:8200`).

Namespace: If you are using HashiCorp Vault Enterprise namespaces, enter the desired namespace.

Authentication

Auth provider: The method for authenticating requests to HashiCorp Vault server. Select one of `Token`, `AWS IAM`, or `AWS EC2`. Your selection determines the remaining **Authentication** options displayed.

Token-based Authentication

Token: Enter the authentication token. This token will be used only to generate child tokens for further authentication actions.

AWS IAM Authentication

In HashiCorp Vault, the term "method" can refer to `userpass`, `token`, or `aws`, among others, but the `aws` method supports two authentication types: `iam` and `ec2`. Meanwhile, in Cribl Edge, you'll see "method" used differently, e.g. in the **Authentication method** setting described below.

Use the **Authentication method** buttons to select one of the following:

- **Auto**: Uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Edge Nodes access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.
- **Manual**: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials directly or by reference. This is useful for Edge Nodes not in an AWS VPC, e.g., those running in a private cloud. It prompts you to provide an **Access key** and a **Secret key**.

Vault AWS IAM Server ID: Value to use for the `Vault-AWS-IAM-Server-ID` header value. This should match the value configured with IAM authentication on Vault.

Vault Role: Authentication role to use in Vault.

Custom auth path: If you enabled authentication in HashiCorp Vault with a custom path, enter that path again here.

For example:

- If you enabled authentication with the HashiCorp Vault command `vault auth enable -path /my-auth aws` instead of `vault auth enable aws` you would set a custom path of `my-auth`. Subsequently, when you perform actions using the `vault write` command, you'd specify an auth type with the `auth_type=ec2` or `auth_type=iam` options.

Assume Role

This section is displayed for all AWS IAM authentication methods.

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION` environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for Vault Auth: Toggle to `Yes` if you want to use your Assume Role credentials to access Vault authentication.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming the role.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

AWS EC2 Authentication

Vault Role: Enter the authentication role to use in Vault.

Custom auth path: If you enabled authentication in HashiCorp Vault with a custom path, enter that path again here. For example:

- You could have used the HashiCorp Vault command `vault auth enable -path /my-auth aws` to enable authentication with a custom path of `my-auth`. Subsequently, when you perform actions using the `vault write` command, you'd specify an auth type with the `auth_type=ec2` or `auth_type=iam` options.

Secret Engine

Mount: Mount point of the Vault secrets engine to use. (Currently, only the KVv2 engine is supported.) Defaults to `secret`.

Secret path: Enter the path on which the Cribl Edge secret should be stored, e.g.: `<somePath>/cribl-secret`.

In a distributed deployment, the Leader, and each Fleet, require a distinct secret. This location cannot be shared between them.

Advanced

Enable health check: Whether to perform a health check before migrating secrets data. Defaults to `Yes`.

Health check endpoint: Configurable endpoint to use for validating system health. Defaults to `/v1/sys/health`.

AWS KMS

Setting the **KMS provider** drop-down to `AWS KMS` exposes the following configuration options:

Authentication

Authentication method: Select an AWS authentication method.

- **Auto**: This default option uses the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`, or the attached IAM role. Works only when running on AWS.
- **Manual**: You must select this option when not running on AWS.

The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to `env.AWS_ACCESS_KEY_ID`, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to `env.AWS_SECRET_ACCESS_KEY`, or to the metadata endpoint for IAM credentials.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION`

environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for KMS: Toggle to `Yes` if you want to use Assume Role credentials to access the AWS KMS.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role. This is required only when assuming a role that requires this ID in order to delegate third-party access. For details, see AWS' documentation.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Service Configuration

KMS Key ARN: Enter the Amazon Resource Name (ARN) of the AWS KMS Key to use for encryption. This entry is required.

When you configure your IAM account/role in AWS, grant access to the following permissions on the KMS key that will be used:

- `kms:Encrypt`
- `kms:Decrypt`

Then use that account for authentication.

9.3. Securing Communications

This page outlines how to protect Cribl Edge Leader to Edge Nodes communications, using an existing TLS/SSL certificate and key.

Cribl Edge expects certificates and keys to be formatted in privacy-enhanced mail (`.pem`) format.
To generate a self-signed certificate and corresponding key, see Securing Cribl Edge.

Importing Certificate and Key

To use your certificate and key to prepare secure communications between Workers/Edge Nodes and the Leader:

1. Navigate to the Leader's **Settings** > **Global Settings** > **Security** > **Certificates** > **New Certificates** modal.
2. Open your TLS/SSL certificate file. (The self-signed certificate example used the placeholder name `myCert.pem`.)
3. Copy the file's contents to your clipboard.
4. Paste the file's contents into the modal's **Certificate** field.

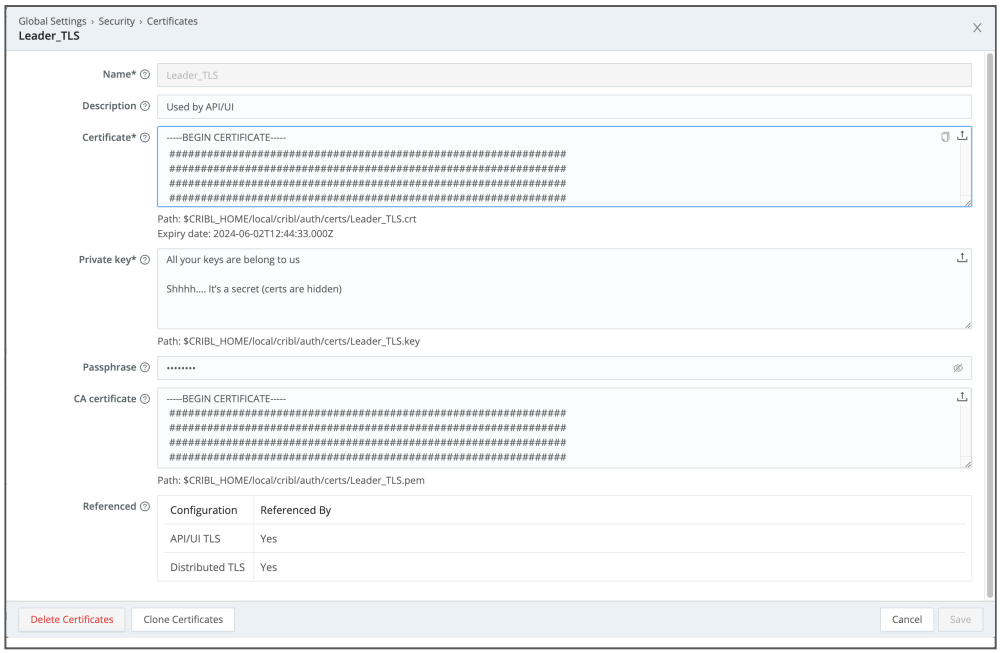
> You can skip the preceding three steps: Just drag/drop your `.pem` file from your filesystem into the **Certificate** field, or upload it using the button at the field's upper right.

5. Open your private key file. (The self-signed certificate example used the placeholder name `myKey.pem`.)
6. Copy the file's contents to your clipboard.
7. Paste the clipboard contents into the same Leader modal's **Private key** field.

> Here again, you can skip the preceding three steps by dragging/dropping or uploading the `.pem` file from your filesystem.

8. Fill in the **Name** and **Description** fields.
9. If you've uploaded a self-signed certificate, just **Save** it now.
10. If your private key is encrypted, fill in the modal's **Passphrase** with the corresponding key. (You can paste the key's contents, or you can drag/drop or upload the key file.)
11. If you're uploading a certificate signed by an external certificate authority – e.g., a downloaded Splunk Cloud certificate – import the chain into the **CA certificate** field before saving the cert. For details, see

Obtain the Certificate Chain (TLS/SSL).



Leader's certificate modal, populated

Connecting to the Leader Securely

You can configure secure communication between your Leader and Edge Nodes using the UI, the `instance.yml` config file, or environment variables.

Using the UI

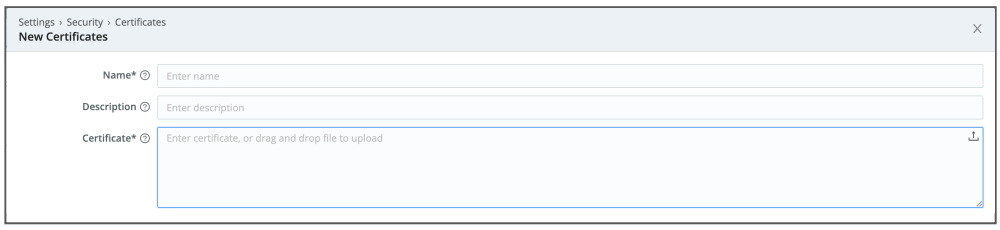
To set up secure communication via the UI, you configure first the Fleet, then the Edge Node, then the Leader.

Fleet Setup

For each Fleet whose Edge Nodes you want to secure:

1. Open your TLS/SSL certificate file, and copy its contents to your clipboard. (This can be the same certificate you uploaded to the Leader, or a different cert.)
2. Select **Manage**, then select the Fleet you want to configure.
3. Select **Group Settings** or **Fleet Settings** (upper right).
4. From the left nav, select **Security** > **Certificates** > **TLS**.
5. Click **Add Certificate**.

6. In the resulting **New Certificates** modal, add the cert's contents to the **Certificate** field.
   (You can paste the cert file's contents, or you can drag/drop or upload the `.pem` file.)
7. As you did on the Leader, also insert your **Private key**, and (as needed) your **Passphrase** and **CA certificate**.
8. Click **Save**.
9. Commit and Deploy the Fleet's new configuration, including the new cert.
10. Repeat the preceding steps on each Fleet.

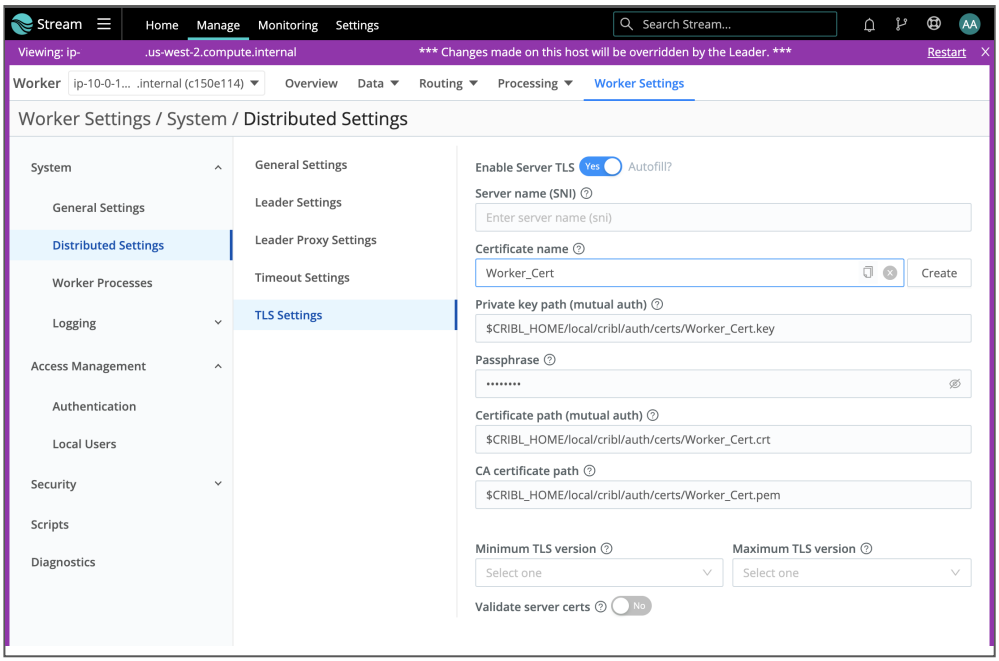


Group-level certificates are configured like the Leader's cert

Edge Node Setup

For each Edge Node that you want to secure:

1. Enable the Leader's UI access to each desired Edge Node (Stream, Edge).
2. Tunnel through from the Leader to a Edge Node's UI.
3. Navigate to this Edge Node's **Worker Settings** (upper right) > **System** > **Distributed Settings** > **TLS Settings**.
4. Toggle **Enable Server TLS** to `Yes`.
   This will expose the remaining TLS settings.
5. From the **Certificate name** drop-down, select the certificate you uploaded to the parent Fleet.
   This will prefill all the required fields. (See all deployed certificates at the left nav's **Security** > **Certificates** link.)
6. Click **Save**.
   The Worker will be unavailable during a short lag, while it restarts with the new configuration.
7. Repeat the preceding steps on each Edge Node.



Configuring TLS on Worker's/Edge Node's UI, from the Leader

Leader Setup

Next, return to the Leader's UI:

1. Select **Settings** > **Global Settings**> **System** > **Distributed Settings** > **TLS Settings**.
2. Toggle **Enable server TLS** to `Yes`.
3. In the **Certificate name** drop-down, select an existing Certificate. This will auto-populate the corresponding cert fields.
4. Click **Save**.

After you've enabled TLS on the Leader, generating bootstrap scripts to add or update Edge Nodes will automatically prepend `https://` to the Leader's URL.

Using YAML Config File

You can also configure the Leader's `$CRIBL_HOME/local/_system/instance.yml` file to ensure that TLS is enabled. Here's the relevant section:

```
distributed:
  mode: managed-edge
  master:
    host: <hostname>
    port: 4200
    authToken: <token>
    tls:
      disabled: false
      rejectUnauthorized: false
      requestCert: false
    resiliency: none
  group: default_fleet
```

After you've enabled TLS on the Leader, generating bootstrap scripts to add or update Edge Nodes will automatically prepend `https://` to the Leader's URL.

Using Environment Variables

Another way to set up secure communications between Edge Nodes and the Leader is via environment variables (Stream, Edge).

If you deploy your Edge Nodes in a container, you can enable encrypted TLS communications with the Leader by configuring the `CRIBL_DIST_MASTER_URL` with the `tls:` protocol. This will override the default setting in `instance.yml`. Here's the format:

```
CRIBL_DIST_LEADER_URL=tls://<authToken>@leader:4200
```

Configuring TLS Mutual Authentication

Once you have configured the Leader for encrypted TLS communication, you can implement client certificate exchange to enable mutual authentication. This allows Cribl Edge to permit only explicitly authorized clients, which hold valid certificates, to connect to the Leader.

When a client certificate is presented to a Leader, two things happen:

1. The Leader validates the client certificate presented to Cribl Edge. The `Validate Client Certs` setting is optional, but highly recommended. When enabled, the Leader checks the certificate against the trust store, to see if it has been signed by a valid certificate authority (CA).

The Leader checks the list of certificates in the `CA Certificates Path` box first (if populated), then against the list of built-in system certificates.

1. The Leader checks whether the `Common Name` (CN) matches the regular expression in the configuration. Cribl Edge's default is to accept any value in the `Common Name` field. You can customize this as needed.

Within the `Common Name`, we validate against the value after the `CN=string`. If your `Common Name` is `CN=logstream.worker`, you would enter `logstream\.worker` in the `Common Name` field – including the backslash, because the value entered is a regular expression.

Limitations on TLS Mutual Auth

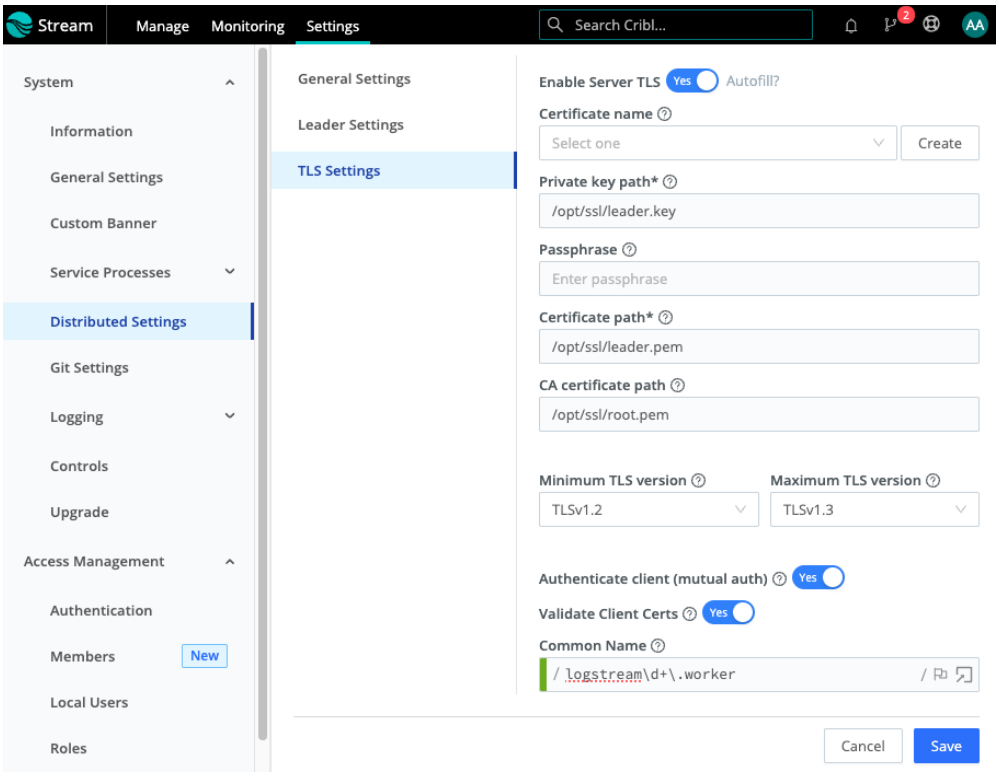
When configuring TLS mutual authentication on Edge Nodes, make sure you place your certificates into a separate directory outside of `$CRIBL_HOME`. If you place the certificates inside `$CRIBL_HOME`, they'll be removed when the next config bundle is deployed from the Leader.

Similarly, you can't bootstrap Edge Nodes with mutual authentication already populated. To bootstrap Edge Nodes, you supply only the shared authentication token. Certificates should be viewed as two-factor authentication; so placing the certificates in the config bundle defeats the purpose of two-factor authentication.

Using the UI

TLS mutual authentication in Cribl.Cloud is configured separately for each Source.

Below is a sample configuration on the Leader:



Configuring Mutual Authentication

In the above configuration, note:

- You can set both the **Minimum TLS version** and **Maximum TLS version**.
- The `Common Name` regex contains an additional check for `\d+`. This allows checking for the format: `logstream1.worker`, `logstream2.worker`, `logstream3.worker`, etc.

Using YAML Config File

To set up TLS **mutual** authentication via the Edge Node's `instance.yml` config file, you need to change some values and also add a few keys, as shown in this example:

```
distributed:
  mode: managed-edge
  master:
    host: <hostname>
    port: 4200
    authToken: <token>
    tls:
      disabled: false
      rejectUnauthorized: true # false if ignoring untrusted certs
      requestCert: true
      privKeyPath: /path/to/certs/worker.key
      certPath: /path/to/certs/worker.pem
      caPath: /path/to/certs/root.pem
    resiliency: none
  group: default_fleet
```

Using Environment Variables

You can set up TLS mutual authentication by configuring this environment variable, using the format shown in this example:

```
CRIBL_DIST_Master_URL="tls://<authToken>@leader.cribl:4200?tls.privKeyPath=/path/t
```

Once you've set this variable, restart the Edge Node. You should see the Edge Node successfully reconnect to the Leader.

If the Edge Node doesn't connect, check `cribl.log` on both the Edge Node and Leader for more context about the problem. You should see errors related to `dist leader communications`.

To build your own Certificate Authority (a self-signed CA), see our blog post on How to Secure Cribl Edge Worker-to-Leader Communications.

TLS Mutual Authentication on Cribl.Cloud

In Cribl.Cloud, you configure TLS mutual authentication separately for each Source.

This requires a CA certificate chain that can validate the client certificate used for authentication. You add your CA certificate by creating a new certificate entry in Cribl.Cloud.

1. Prepare the CA certificate chain PEM file.
2. Go to a Edge Node's **Settings** > **Security** > **Certificates** and select **Add Certificate**.
3. Populate the **Certificate** field with any valid PEM-formatted content:

```
-----BEGIN CERTIFICATE-----
CERTIFICATE CONTENT
-----END CERTIFICATE-----
```

   The certificate and key are required only for UI validation and are not used otherwise.
4. Populate the **Private key** with the key in PEM format:

```
-----BEGIN RSA PRIVATE KEY-----
HIDDEN PRIVATE KEY
-----END RSA PRIVATE KEY-----
```

5. In the **CA certificate** field, enter your PEM-formatted certificate and save. This will generate the CA certificate path.
6. Edit the certificate again to view the certificate path and copy it.
7. Go to the Source where you want to enable mutual TLS and paste the path in the **CA certificate path** field.
8. Save, commit, and deploy to finish the process.

Setting Authentication on Sources/Destinations

You can use certificates to authenticate Cribl Edge to external data senders and receivers. You configure this at the Group level, as follows:

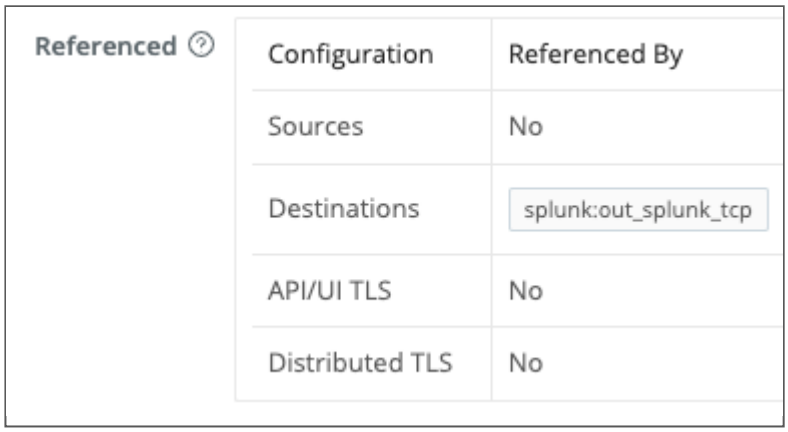
1. Select a Group.

As an alternative to the preconfiguration in steps 2–6, you can import a certificate on the fly, using the Source's or Destination's **Create** button. See this section's final steps below.

2. Open **Group Settings** (top right) > **Security** > **Certificates**.
3. Select **New Certificates**.
4. Paste or upload `.pem` files, as in Setting Up the Encrypted Channel.
5. Supply a **Passphrase** and/or **CA certificate**, if required by your integration partner's certificate.
6. Click **Save**.

7. Open your relevant Source's or Destination's config modal.
8. Select **TLS Settings (Client Side)** or **TLS Settings (Server Side)**, depending on the integration.
9. Slide **Enabled** on.
10. In the **Certificate name** drop-down, select a cert that you've preconfigured for this integration. This will auto-populate the corresponding fields here.
11. If you're creating a certificate on the fly, click the **Create** button beside **Certificate name**.
12. Click **Save**.

Cribl Edge will create new certificates at the same Group level that you're configuring. You can verify this at the **Create new certificate** modal's bottom, by making that the **Referenced** table includes rows for poulated for the appropriate **Sources** and **Destinations**.



Group-level certificate modal

Securing the Leader Node

This is a best practice that enables the Leader to validate itself to clients. We can secure it using the self-signed cert we created in Securing Cribl Edge:

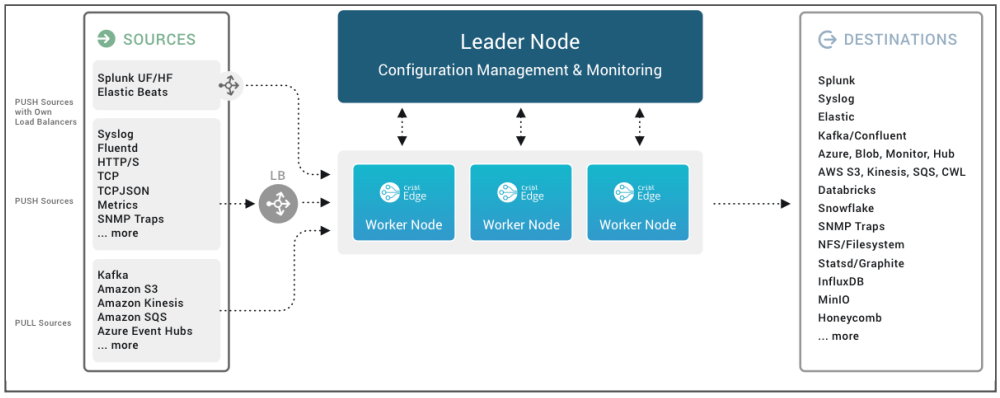
1. Navigate to **Settings** > **Global Settings** > **General Settings** > **API Server Settings** > **TLS**.
2. Slide the toggle to **Enabled**.
3. From the **Certificate Name** drop-down, select a cert you've previously imported. This will populate the corresponding fields here.
4. Click **Save**.

After this save, you must prepend `https://` to all Cribl Edge URLs on the Leader Node. E.g., to get back to the Settings page you just configured, you'll now need to use `https://<hostname>:<port>/settings/system)`.

10. Sources

Each Cribl Edge **Source** is a configuration that enables Edge nodes to collect or receive observability data – logs, metrics, application data, etc. – in real time. Edge can receive continuous data input from Splunk, HTTP senders, Elastic Beats, Prometheus, TCP JSON, and many others. Sources can receive data from either IPv4 or IPv6 addresses.

Edge's UI offers a configuration modal for each type of supported Source. However, you can add multiple instances of each Source type – with each configured to match the parameters of the corresponding sender. E.g., you can have multiple File Monitors and multiple listeners for Syslog, Splunk, Elastic Beats, Prometheus, TCP JSON, and many others.



Sources in the Edge ecosystem

System and Internal Sources

Sources that generate data locally at the Edge Node; or monitor resources; or move data among Edge Nodes and/or Stream Workers within your Cribl deployment.

- AppScope
- Cribl Internal
- Cribl HTTP
- Cribl TCP
- Datagen
- Exec
- File Monitor
- Kubernetes Logs
- Kubernetes Metrics
- System Metrics

- System State
- Windows Metrics

PUSH Sources

Supported data Sources that Cribl Edge **fetches** data from.

These Sources can continue ingesting data even if no Leader is active:

- Amazon Kinesis Firehose
- Datadog Agent
- Elasticsearch API
- Grafana
- HTTP/S (Bulk API)
- Loki
- Metrics
- Raw HTTP/S
- OpenTelemetry (OTel)
- Prometheus Remote Write
- SNMP Trap
- Splunk HEC
- Splunk TCP
- Syslog
- TCP JSON
- TCP (Raw)
- UDP (Raw)
- Windows Event Forwarder

PULL Sources

Supported data Sources that Cribl Edge **fetches** data from.

These Sources can ingest data only if a Leader is active:

- Prometheus Edge Scraper
- Prometheus Scraper (Deprecated)
- Windows Event Logs

Configuring and Managing Sources

For each Source *type*, you can create multiple definitions, depending on your requirements.

To configure Sources, from the top nav, click **Manage**, then select a **Fleet** to configure. Then, you have two options:

- To access the graphical QuickConnect UI, click **Collect**. Next, click either **Add New** or (if displayed) Select **Existing**.
- To access the Routing UI, click **More** > **Sources**. On the resulting **Data Sources** page's tiles or left menu, select the desired type, then click **Add New**.

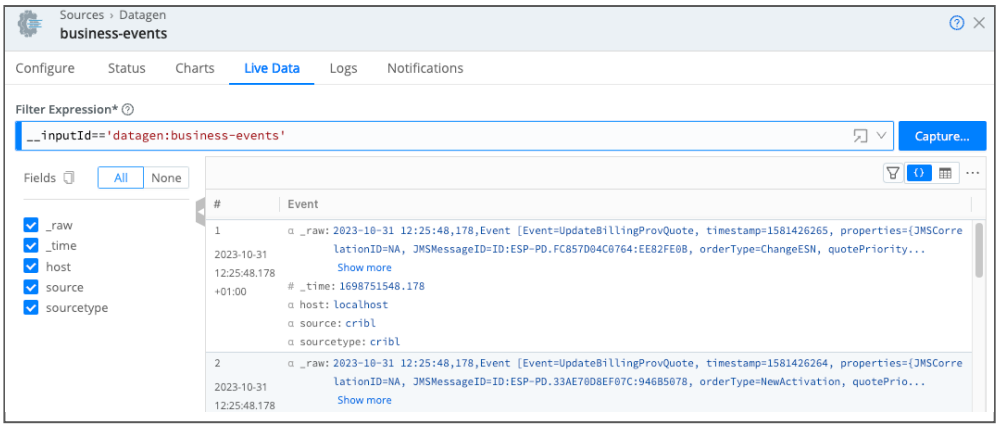
Capturing Source Data

To capture data from a single enabled Source, you can bypass the Preview pane, and instead capture directly from a **Manage Sources** page. Just click the **Live** button beside the Source you want to capture.

In order to capture live data, you must have Edge Nodes registered to the Fleet for which you're viewing events. You can view registered Edge Nodes from the **Status** tab in the Source.

Source > Live button

You can also start an immediate capture from within an enabled Source's configuration modal, by clicking the modal's **Live Data** tab.



Source modal > Live Data tab

Monitoring Source Status

Each Source's configuration modal offers two tabs for monitoring: **Status** and **Charts**.

Status Tab

The **Status** tab provides details about the Edge Nodes in the Fleet and their status. An icon shows whether the Edge Node is operating normally.

You can click each Edge Node's row to see specific information, for example, to identify issues when the Source displays an error. The specific set of information provided depends on the Source type. The data represents only process 0 for each Edge Node.

The content of the **Status** tab is loaded live when you open it and only displayed when all the data is ready. With a lot of busy Edge Nodes in a group, or nodes located far from the Leader, there may be a delay before you see any information.

The statistics presented are reset when the Edge Node restarts.

Charts Tab

The **Charts** tab presents a visualization of the recent activity on the Source. The following data is available:

- Events in
- Thruput in (events per second)
- Bytes in
- Thruput in (bytes per second)

This data (in contrast with the status tab) is read almost instantly and does not reset when restarting an Edge Node.

Preconfigured Sources

To accelerate your setup, Cribl Edge ships with several common Sources configured but not switched on. Open, clone (if desired), modify, and enable any of these preconfigured Sources to get started quickly:

- **System Metrics** – Basic Level
- **File Monitor** > **in_file_auto** – Auto Discovery Mode
- **File Monitor** > **in_file_varlog** – Manual Discovery Mode
- **AppScope** > **in_appscope** – Unix Domain Socket listener
- **Cribl Internal** > **CriblLogs** – Internal
- **Cribl Internal** > **CriblMetrics** – Internal

Cribl University offers a course titled Collecting Data in Edge that provides an illustrated overview. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

10.1. System

10.1.1. Exec

The Exec Source enables you to periodically execute a command and collect its `stdout` output. This is typically used in cases when Cribl Edge cannot accomplish collection with native Collectors or other Sources.

Available in Cribl.Cloud: **No** | Type: **System** | TLS Support: **N/A** | Event Breaker Support: **Yes**

This Source allows you to run **almost anything** on the host system. Make sure you understand the security and other impacts of commands you plan to execute, before proceeding.

Especially for monitoring or polling, you'll need to receive the command or script's output periodically. For this reason, the Exec Source lets you specify when the command or script should run, by time interval or by cron-style schedule.

Here are a few examples of what you can run from an Exec Source:

- Database queries.
- Custom commands to poll databases or apps.
- `ping` to check latency.
- `ps -ax` to get a list of running processes.
- `SNMP GET` requests to query an SNMP agent about network entities.
- `netstat -natp` to get lists of sockets and TCP ports, and what they're doing.
- `mtr -c 1 --report --json 8.8.8.8` to run a traceroute on a location (here, `8.8.8.8`), and packaging up a report in JSON format.

Configuring an Exec Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Exec**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **Exec**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Enabled: Defaults to `Yes`.

Input ID: Enter a unique name to identify this Exec Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Command: Command to execute; supports Bourne shell syntax.

Schedule type: Use the buttons to select either `Interval` or `Cron`. Customize the behavior in the corresponding field below the button.

- **Interval**: Specify how often, in seconds, the command should run. Defaults to `60`.
- **Schedule**: Enter a cron expression. (You enter the cron expression in UTC time, but the resulting **Estimated Schedule** displays in local time.)

Optional Settings

Max retries: Maximum number of retry attempts in the event that the command fails.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

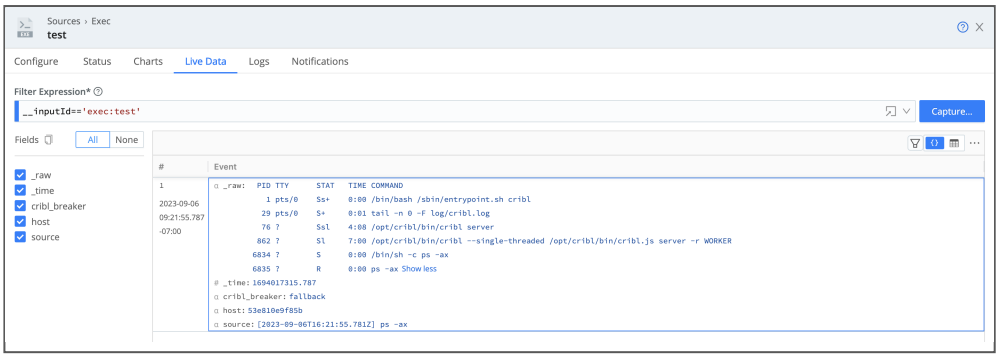
Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses internal fields to assist in forwarding data to a Destination.

You can find the following event fields on the **Live Data** tab of the selected Exec Source, including the hostname and source of the event:

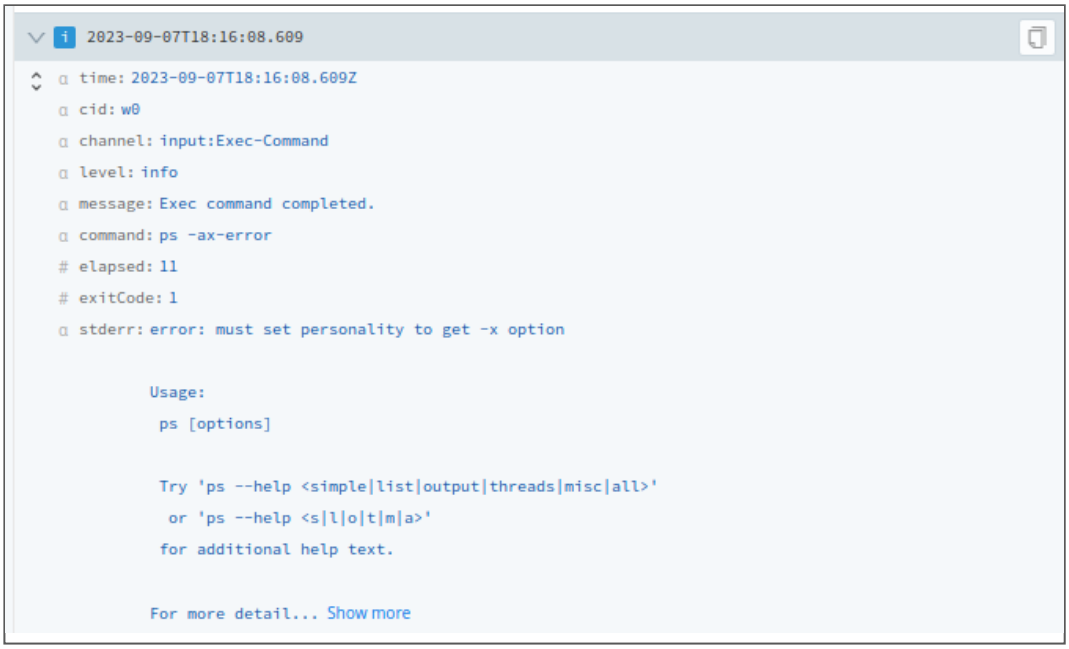
- `_raw`
- `_time`
- `cribl_breaker`
- `host`
- `source`



Event fields

Troubleshooting

If a command isn't running as expected, the **Logs** tab can help you identify issues by displaying the command executed, its elapsed time, and its exit code.



Event fields

10.1.2. File Monitor

The File Monitor Source generates events from text files, compressed, archived, and (some) binary log files, based on lines and records extracted from the content.

Available in Cribl.Cloud: **No** | Type: **System** | TLS Support: **N/A** | Event Breaker Support: **Yes**

As of v.4.2.x, this Source can process the following file formats:

- **Compressed/archived**: `zip`, `gzip`, `zstd`, and `tar`. Cribl Edge only supports DEFLATE compression for `zip` files.
- **ASCII**: `text`.
- **Non-ASCII binary**.

Binary files are broken into base64-encoded chunks and streamed bypassing Event Breakers. Text files are processed normally through Event Breakers. (Before v.4.2.x, the File Monitor Source could handle only text files.) Also as of v.4.2.x, the File Monitor Source no longer excludes `*.gz` files by default.

Discovering and Filtering Files to Monitor

To produce its initial list of files to monitor, the File Monitor Source runs a discovery procedure at a configurable **Polling interval**. The Source then applies an **Allowed list** to filter the initial list down into its final form. Then, for each file on the list, the Source compares current state with previously-stored state. This comparison determines whether the File Monitor Source will actually watch a given file for a given polling interval, or just ignore the file.

In the simplest case, the Source discovers a file for which it has no stored state. This means that the file has just been created, and needs to be monitored. See the Examples for other possibilities, along with a description of the **Status** tab, which displays state information for all files being monitored.

How does the File Monitor Source discover files in the first place? You have the choice of two **Discovery Modes**: **Auto** or **Manual**.

- In **Auto** mode, the Source automatically discovers files that running processes have open for writing. Auto mode collects logs from (the discovered) active files that match the path/allowlist.
  Auto mode is useful to detect which files are being written to. For example, if you enter an allowlist of `*log`, Auto mode will find all the active logs within that directory, no matter what's writing to it, and it will exclude any rotated logs.

Cribl Edge currently supports Auto mode only when running on Linux, not on Windows.

- In **Manual** mode, the Source discovers files based on the specified directory and depth. **Manual** mode is often used to collect logs in a folder where the process that creates them rotates them when they get large. For example, if you specify a path of `/var/log` and an allowlist of `*/messages*`, Manual mode will capture the current `/var/log/messages`, as well as older logs from `/var/log/messages*`.

If you are using a tool like `rsync` to transfer files in chunks, the File Monitor Source might start collecting files before the transfer is complete. To prevent this, configure your tool to transfer files serially. (For example, use: `rsync --append`.)

Configuring a File Monitor Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

Configure via QuickConnect

1. To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge).
2. Click **Add Source** at left. From the resulting drawer's tiles, select **System and Internal** > **File Monitor**.
3. Click either **Add Destination** or (if displayed) **Select Existing**.

The **General Settings** drawer will open.

Configure via Routing

1. Click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge).
2. From the resulting page's tiles or left nav, select **System and Internal** > **File Monitor**.
3. Click **Add Source** to open the **New Source** modal.

General Settings

Enabled: Toggle to `Yes` to enable the Source.

Input ID: Enter a unique name to identify this File Monitor Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Discovery Mode: Use the buttons to select one of these options:

- **Auto**: Tells the Source to automatically discover files that running processes have open for writing.
- **Manual**: Tells the Source to discover the files within the **Search path** (i.e., a directory) that you specify, down to the **Max depth**. The defaults are:
    - **Path** is set to the `/var/log` directory
    - **Allowlist** of `*/log/*` and `*log`
    - **Max depth** of 4

If you leave the **Max depth** field empty, the Source will search subdirectories, and their subdirectories, and so on, without limit. If you specify `0`, the Source will discover only the top-level files within the **Search path**. For `1`, the Source will discover files one level down.

Both modes allow you to set the **Polling interval**, which otherwise defaults to 10 seconds.

Optional Settings

Filename allowlist: Wildcard syntax, and the exclamation mark (`!`) for negation, are allowed. For example, you can use `!*cribl*access.log` to prevent the Source from discovering Cribl Edge's own log files. The default filters are `*/log/*` and `*log`.

Always think through how your **Search path** and **Filename allowlist** settings interact: If you're not careful, you can inadvertently create overlapping monitor groups that deliver duplicates of some files. See Using Allowlists Effectively below.

Max age duration: Optionally, specify a maximum age of files to monitor. This Source will filter events with timestamps newer than the configured duration. (Where files don't have a parsable timestamp, it will set their events' timestamp to `Now`.) Enter a duration in a format like `30s`, `4h`, `3d`, or `1w`. Defaults to an empty field, which applies no age filters.

Check file modification times: If toggled to `Yes`, this Source will skip files with modification times older than the configured **Max age duration**.

When you set a **Max age duration** threshold, the Source will open every file and read through all its contents to seek a timestamp. For best performance, also enable the **Check file modification times** toggle to skip older files.

Collect from end: If toggled to `Yes`, then upon Cribl Edge's next startup, this Source will skip to the end of new files. (It will run discovery once, adding any newly discovered files to the state store as if they had already been consumed.) After that initial run, when this Source next discovers new files, it will read those new files from the head.

Enable binary files: If toggled to `Yes`, the Source will report the file as binary and stream it in Base64-encoded chunks. By default, the Source ignores binary (non-text) files such as JPEG images, MP3 audio files, or some binary data files.

Force text format: If toggled to `Yes`, the Source displays ingested data as text in the event (when the data isn't in a compressed or archived file). This option is helpful if you have files that contain binary mixed with text where the data should be streamed as all text (for example, AuditD logs can contain this type of data).

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec.), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Idle timeout: Time, in seconds, before an idle file is closed. Defaults to `300` sec. (5 minutes).

Hash length: How many file header bytes to use in a hash for identifying a file uniquely. Defaults to `256` bytes. For details, see Configuring Hash Lengths.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

This Source defaults to QuickConnect.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are not part of an event, but they are accessible, and Functions can use them to make processing decisions.

For the File Monitor Source, you can use internal fields to enrich events with container and host metadata. The following internal fields are available:

- `__baseFilename`
- `__source` / `source`
- `__raw`
- `__inputId`
- `container_id`
- `container_path`
- `host_path`

Monitoring Renamed Files' State

The File Monitor Source uses a simple scheme to find "backlog" files corresponding to a matching file: Tell it to tail `foo.log`, and it will look for `foo.log.[0-9]` and scrape those, too.

This Source keeps hashes that correspond to the start point within the file, and to the last-read point. This way, if Cribl Edge is stopped, files are rotated, and Cribl Edge is restarted, the File Monitor Source can find

the resume point in those backlog files.

Similarly, if you rename the file **within the same backlog scheme** (e.g., `foo.log` to `foo.log.0`), the File Monitor Source can resume at the right place. However, if you renamed `foo.log` to `bar.log` (a deviation from the expected naming scheme), this Source could not find its resume point.

Configuring Hash Lengths

For files with identical first lines, configure the hash length to be longer than the first line to ensure that the header includes something unique. It is common for a batch of `CSV` files or `IIS` logs to have identical first lines listing columns that appear in subsequent lines.

If you don't adjust the hash length, all files with headers larger than `256 bytes` will appear to File Monitor as one file. As long as the hash length is greater than the header length, and the subsequent lines include something unique, the File Monitor Source will identify them as different files.

Using Allowlists Effectively

Done properly, file monitoring collects exactly the files you need to see, without duplication. You want to avoid inadvertently creating overlapping monitor groups that deliver duplicates, because this can "run the meter" twice for affected files, potentially causing license issues.

The key is knowing how to craft allowlists.

Crafting Allowlists: Principles

To get the results you want from file monitoring, always apply the following principles.

Write allowlists to explicitly include the full path of files you want to collect.

The search path itself is **not** excluded from the match. For example, if you want to collect the file `/var/log/messages` when you've set your search path to `/var/log`, your filename allowlist needs to be either `/var/log/messages`, `/var/*/messages`, or `*/log/messages`.

Exclusions (`!`) also need to address the full file path. For example, if you want to collect everything from `/var/log/*` **except for** `/var/log/apache/*`, your allowlist must include one of the following:

- `!/var/log/apache/*`
- `!*/apache/*`
- `!*/log/apache/*`

Wildcards (`*`) match **any** part of the file path:

- `*log` matches `/var/foo/bar/myfile.log`.
- `*.log` matches anything ending in `.log`.
- `/var/log/*` matches anything in `/var/log`, subject to the **Max Depth** setting.

Put explicit matches (and exclusions) at the beginning of the allowlist, ahead of wildcard matches.

For example, an allowlist of `/var/log/*` followed by `!/var/log/apache/*` will fail to exclude the file `/var/log/apache/foo` because that file matches the first allowlist entry.

In allowlists, order matters!

Be as specific as possible, whenever possible, to prevent accidental matches.

Each of the following examples shows how to avoid matching unwanted files within the directories that you're monitoring.

To collect `/var/log/boot.log` (and older versions of `boot.log.*`) as well as `/var/log/messages`:

Search path	Filename allowlist
`/var/log`	`/var/log/boot.log`, `/var/log/messages`

To collect everything in `/var/log` while excluding `/var/log/apache/*`:

Search path	Filename allowlist
`/var/log`	`!/var/log/apache/*`, `/var/log/*`

Example: Excluding the System's Own Logs

Recall that you'll often create a set of File Monitor Sources, each of which monitors one file. This is **not** the right approach in an example like this one, where **Discovery Mode** is set to **Auto**. Using **Auto** mode in a file monitor along with literally any other file monitor can cause unintended, duplicate collection. Use **Auto** mode carefully!

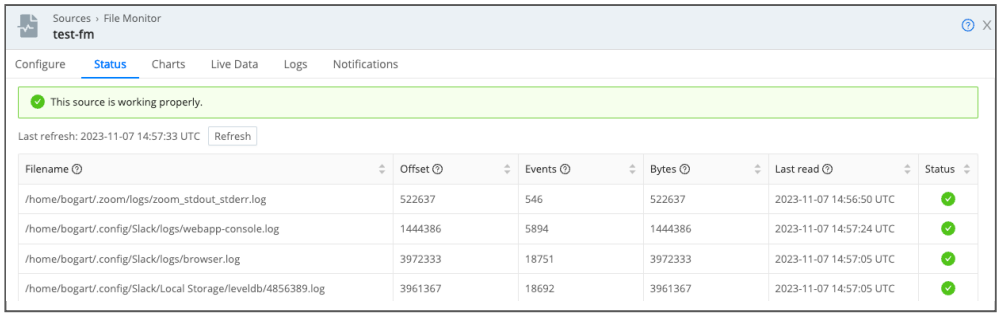
Here's the example:

Suppose you added a File Monitor Source on a Linux machine, set your **Discovery Mode** to **Auto**, and specified your home directory as the **Search path** (e.g., `/home/bogart/`).

You do not want to monitor Cribl Edge's own log files, or any log files generated by Chrome, the web browser you're using. To exclude them, you add `!*cribl*access.log` and `!*chrome*` to your **Allowlist**. You **do**

want to monitor any other log files that the Source can discover, so you also add `*log`.

After awhile, the Status tab looks like this:



The Status tab

Example: Fixing Overlapping Monitor Groups

Here's an example of the kind of overlapping monitor groups you should avoid:

Search path	Max depth	Filename allowlist
`/var`	(none)	`*/log/*`
`/var/log/apache`	`0`	`*`

The above configuration would collect all the files in `/var/log/apache/` twice. A better solution would be the following:

Search path	Max depth	Filename allowlist
`/var/log`	(none)	`!*/apache/*`
`/var/log/apache`	`0`	`*`

10.1.3. Journal Files

The Journal Files Source collects data from systemd's centralized logging, which is called the journald service. This service is a centralized location for all messages logged by different components in a systemd-enabled Linux system. This includes messages from kernel, boot, syslog, or other services. This Source is currently configured to configured to collect data from the `user` journal, as `system` is privileged. This Source is disabled by default.

Type: **System** | TLS Support: **N/A** | Event Breaker Support: **No**

The Journal Files Source is available on Cribl.Cloud hybrid Workers, but not on Cribl-managed Cribl.Cloud Workers.

Discovering and Filtering Journal Files to Monitor

To determine all the names of all journals to read, the Journal Files Source runs a discovery procedure at a configurable **Polling interval**. The Source then applies an **Allowed list** to filter the initial list down into its final form. Then, for each journal on the list, the Source compares current state with previously stored state. This comparison determines whether the Journal Files Source will actually stream a given journal for a given polling interval. The Source then runs the files through the **Filter Rules** which allows you to set additional criteria.

Cribl Edge ships with a single, disabled instance of the Journal Files Source with a default configuration. The search path is configured to point to `$CRIBL_EDGE_FS_ROOT/var/log/journal/$MACHINE_ID` and defaults to streaming the primary `system.journal` it finds there. `$MACHINE_ID` is an environment variable that you can insert into the Source's search path. If you use the environment variable but don't assign a value, the Source will look up the local host `machine id` from `/etc/machine-id` and use that value.

Currently, this Source does not support all of the compression options that the journald service supports. The parser this Source uses to read the files can handle `LZ4` and `ZSTD` compression. When the parser encounters compressed fields it can't unpack, an error message emits in the logs. Below is an example of the error message:

```
Error reading journald event, dropping offset=31678152 evt="Mar 02 21:29:54
goats-xps15 multipassd: undefined"
```

> Also, this Source does not support the new `COMPACT` binary file format. The logs will contain an error message when the parser encounters this type of file.

Configuring a Journal Files Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Journal Files**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **Journal Files**. Next, click **Add Source** to open a **New Source** modal that provides the options below.

General Settings

Enabled: Toggle to `Yes` to enable the Source.

Input ID: Enter a unique name to identify this File Monitor Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Search path: Specify the directory to search journals on. Accepts environment variables. The default value for this field varies depending on the product:

- In Cribl Edge, it defaults to `$CRIBL_EDGE_FS_ROOT/var/log/journal/$MACHINE_ID`.
- In Cribl Stream, it defaults to `/var/log/journal/$MACHINE_ID`.

Journal allowlist: The full path of each discovered journal is matched against this wildcard list. Defaults to `system`.

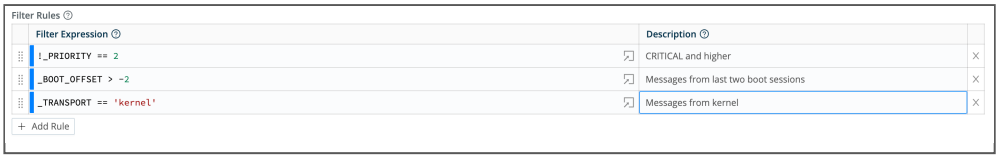
Optional Settings

Polling interval: How often, in seconds, to collect metrics. If not specified, defaults to `10` seconds.

Filter Rules: Optionally, specify which journal objects to allow. Events are generated if all the rules' expressions evaluate to true, or if no rules are specified.

- **Filter expression**: JavaScript expression to filter Journal objects. Returns `true` to include it.
- **Description**: Optional description of the rule.

The default **Filter expression** is `severity <= 4` with a **Description** of `Allow events having 'emergency', 'alert', 'critical', 'error', or 'warning' priority.`
The screenshot below shows other examples you can use:



Filter expressions

Current boot only: Skip events that are not part of the current boot session.

Max age duration: Optionally, specify a maximum age of journal objects to stream. This Source will filter events with timestamps newer than the configured duration. Enter a duration in a format like `30s`, `4h`, `3d`, or `1w`. Defaults to an empty field, which applies no age filters.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

In Cribl Edge, this Source defaults to QuickConnect. While, in Cribl Stream, it defaults to Send to Routes.

10.1.4. Kubernetes Events

The Kubernetes Events Source relies on the Kubernetes API watch capabilities to collect cluster-level events. These events are generated automatically in response to state changes or errors with nodes, pods, or containers. The Kubernetes API server acts as a notification server for the events we are watching in a given Kubernetes cluster. This Source acts in a listener/subscriber capacity to collect events for all namespaces from the cluster API. For a sample of the events captured, see Live Data.

To collect cluster-level events, the Kubernetes Events Source needs watch access to the Kubernetes API. For details, see RBAC for Kubernetes Events Source.

Type: **System and Internal** | TLS Support: **N/A** | Event Breaker Support: **No**

Cribl Edge supports configuring only one Kubernetes Events Source per Edge Node and per Fleet.

On Cribl.Cloud, this Source is available on hybrid Workers, but Cribl-managed Workers omit this Source – Cribl manages these Workers' uptime and diagnostics on your behalf.

Discovering and Filtering Kubernetes Events

The Kubernetes Events Source connects to the Kubernetes API to set up a watch, and to request a starting point from the latest event at 5-minute polling intervals. The Source then runs the lists through the **Filter Rules** to determine what to report on.

This Source emits events in a way that might include some redundancies in the properties extracted. For details on the output, plus guidelines on what fields to drop, see Event Structure Details.

If no rules are configured, or if all of the rules evaluate to `true`, this Source generates events for the object. Conversely, if any of the rules evaluates to `false`, this Source skips the object, and does not generate any events for it.

Cluster/Daemonset Best Practices

For the Kubernetes Events Source to work as designed, Cribl recommends deploying Cribl Edge as a Daemonset. For details, see Deploying via Kubernetes.

This Source can generate redundant events when it is running inside the cluster, and can't determine which DaemonSet it's in or which Node it's on. This happens when you run Cribl Edge inside the cluster, but in a

way that it can't detect the local Pod/node. In these cases, Edge will emit an error when it initializes this Source.

You can bypass the error state by setting the `CRIBL_K8S_FOOTGUN` environment variable to `true`. However, you will assume the risk that the Kubernetes Events Source might make excessive calls to the cluster API.

Configuring Cribl Edge to Collect Kubernetes Events

From the top nav, click **Manage**, then select a **Fleet** to configure. Then, you have two options:

- To configure via the graphical QuickConnect UI, click **Collect**. By default, a **Kubernetes Events** tile appears at left. Hover over it and select **Configure** to open a drawer that provides the options below.
- To configure via the Routing UI, click **More** > **Sources**. From the resulting page's tiles or the **Sources** left nav, select **System and Internal** > **Kubernetes Events**. Next, click the default `in_kube_events` Source to open a modal that provides the options below.

General Settings

Input ID: This is prefilled with the default value `in_kube_events`, which cannot be changed via the UI, due to the single-Source restrictions above.

Optional Settings

Filter Rules: Optionally, specify the Kubernetes objects which this Source should parse to generate events. If you specify no restrictive rules here, the Source will emit all events.

- **Filter expression**: JavaScript expression to filter Event objects. Filters are based on the Event Objects, and evaluated from top to bottom. The first expression evaluating to `false` excludes the Pod from collection. The default filter is: `!metadata.namespace.startsWith('kube-')`, which ignores Pods in the `kube-*` namespace.
  Other **Filter expressions** examples include:
  - Ignores events in the `kube-*` namespaces – `!object.metadata.namespace.startWith('kube-')`
  - Ignore all events from DaemonSet – `object.regarding.kind != 'DaemonSet'`
  - Collects events matching a specific object name - `object.regarding.name == 'object name'`

- **Description**: Optional description of the rule.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

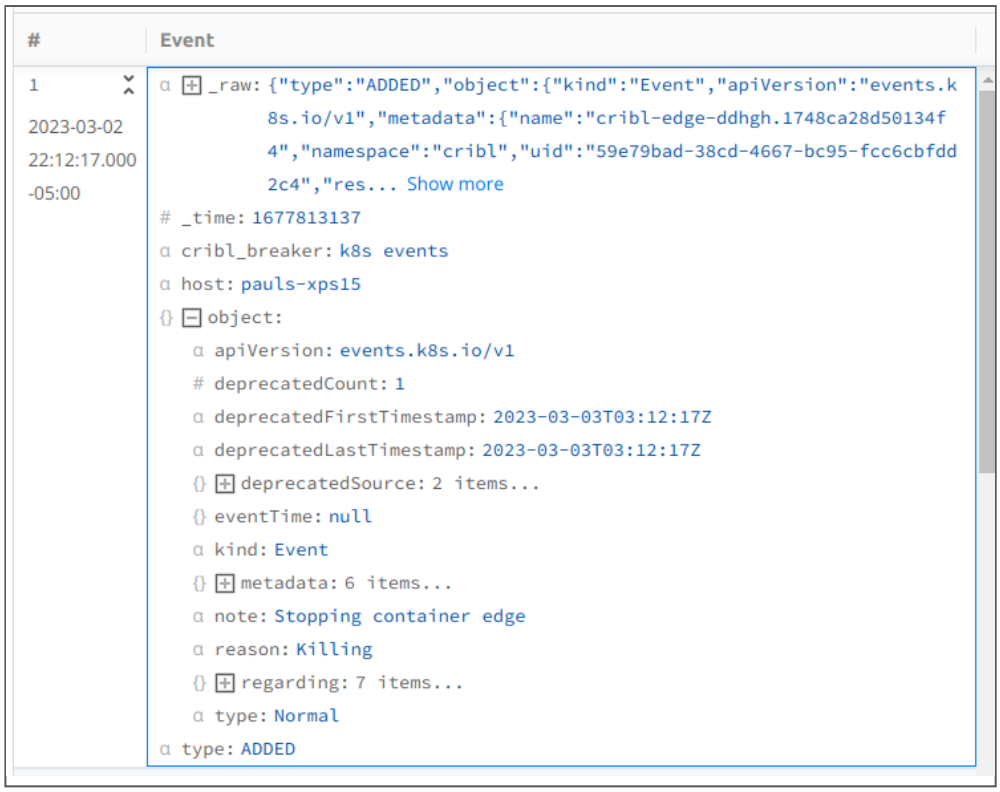
Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Event Structure Details

This Source emits a `_raw` property as a JSON-encoded object, as other Sources do. However, there are additional `object` and `type` properties from that extraction. The logic behind this is that `timestamp` and other fields had to be extracted to keep track of starting points.

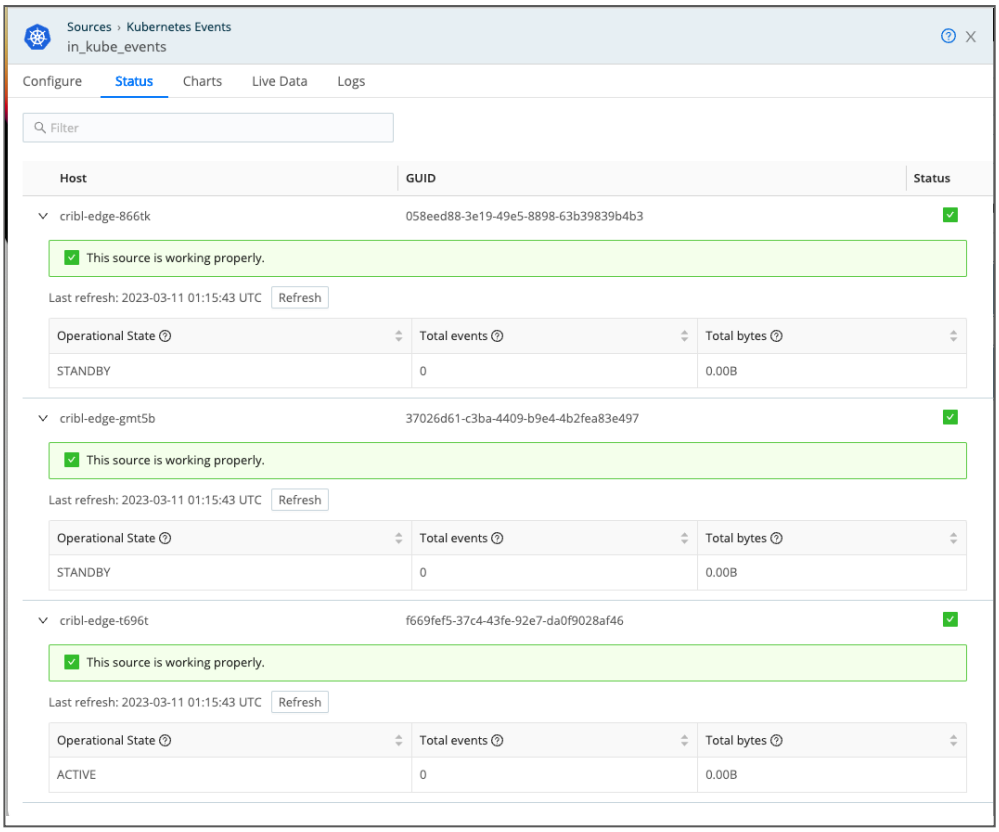
If you are sending the events over a metered connection, then you will get three copies of the same data: `_raw`, `__raw`, and the extracted `object` fields. Consider dropping the redundant fields before sending them to your Destination(s).



Sample Event

Source Operational State

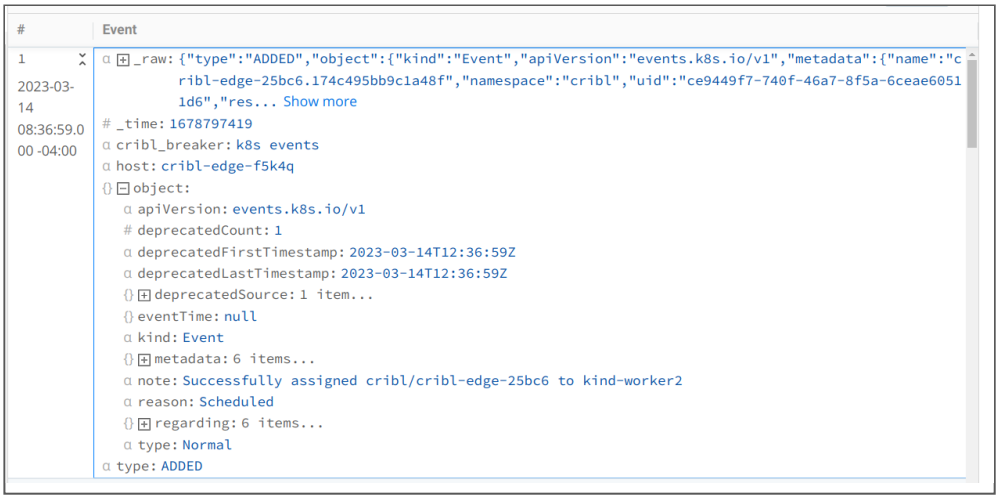
To check the Source's operational state, go to **Status** and expand the host details. The **Operational State** Column shows either an **Active** state or **Standby**. **Active** indicates the Source is running or won the election. **Standby** means it's waiting to be re-elected and not currently running.



Operational State

Source Live Data

To see a sample of the Source's data, click the **Live Data** tab.



Sample Event

10.1.5. Kubernetes Logs

The Kubernetes Logs Source collects logs from containers on a Kubernetes node. Optionally, you can filter and enrich incoming logs. You must authorize this Source to access the Pods in all namespaces. For details, see RBAC for Kubernetes Logs Source.

Type: **System and Internal** | TLS Support: **N/A** | Event Breaker Support: **Yes**

Cribl Edge supports configuring only one Kubernetes Logs Source per Edge Node and per Fleet. This Source is currently unavailable on Cribl-managed Cribl.Cloud Workers.

Discovering and Filtering Kubernetes Pods to Monitor

The Kubernetes Logs Source connects to the Kubernetes API and loads the lists of Pods on the node, on a configurable **Polling interval**. The Source then runs the Pods through the **Filter Rules** to determine which ones to report on.

If no rules are configured, or if all of the rules evaluate to `true`, the Source generates logs for the Pod's containers. Conversely, if any of the rules evaluate to `false`, the Source skips the Pod's containers, and does not generate any logs for it.

Cluster/Daemonset Best Practices

For the Kubernetes Logs Source to work as designed, Cribl recommends deploying Cribl Edge as a Daemonset. For details, see Deploying via Kubernetes.

This Source will refuse to run when it connects to the cluster API instead of to the `kubelet` on the local node. This happens when you run Cribl Edge outside of the cluster; or when you deploy it inside the cluster, but in a way that it can't detect the local Pod/node. In these cases, Edge will emit an error when it initializes this Source.

You can bypass the error state by setting the `CRIBL_K8S_FOOTGUN` environment variable to `true`. However, you will assume the risk that the Kubernetes Logs Source might make excessive calls to the cluster API.

Configuring Cribl Edge to Collect Kubernetes Logs

From the top nav, click **Manage**, then select a **Fleet** to configure. Then, you have two options:

- To configure via the graphical QuickConnect UI, click **Collect**. By default, a **Kubernetes Logs** tile appears at left. Hover over it and select **Configure** to open a drawer that provides the options below.
- To configure via the Routing UI, click **More** > **Sources**. From the resulting page's tiles or the **Sources** left nav, select **System and Internal** > **Kubernetes Logs**. Next, click the default `in_kube_logs` Source to open a modal that provides the options below.

General Settings

Input ID: This is prefilled with the default value `in_kube_logs`, which cannot be changed via the UI, due to the single-Source restrictions above.

Optional Settings

Polling interval: How often, in seconds, to collect metrics. If not specified, defaults to `15` seconds.

Filter Rules: Optionally, specify the Kubernetes Pods that this Source should parse to generate logs. If you specify no restrictive rules here, the Source will emit all events.

- **Filter expression**: JavaScript expression to filter Kubernetes Pod objects. Filters are based on the Kubernetes Pod Object definition, and evaluated from top to bottom. The first expression evaluating to `false` excludes the Pod from collection. The default filter is `!metadata.namespace.startsWith('kube-')`, which ignores Pods in the kube-* namespace.

Other **Filter expressions** examples include:

- Collect logs from Pods on a specific Node – `spec.nodeName == 'node1'`
- Ignore all DaemonSets – `metadata.ownerReferences[0].kind != 'DaemonSet'`
- Ignore Pods with specific Container names – `spec.containers[0].name != 'edge'`
- **Description**: Optional description of the rule.

Pods are either included or excluded entirely based on the Filter Rules.

Enable timestamps: When toggled to `Yes`, Cribl Edge prefixes a timestamp to each line of the container's raw console output. When you enable timestamps, you must select the `kubernetes_logs` pre-processing Pipeline. This Pipeline removes conflicting timestamps prepended by the Source. This combination is

designed for working with containers whose console output lacks proper timestamps (either because the timestamps are missing altogether or because they specify time only but not date).

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Event Breakers

This section defines event breaking rulesets that will be applied, in order, on the `/raw` endpoint.

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec.), maximum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

Select a **Pipeline** (or Pack) from the drop-down to process this Source's data. Required to configure this Source via **Data Routes**; optional to configure via **Collect**/**QuickConnect**.

Disk Spooling

Enable disk spooling to search for spooled Kubernetes logs directly from Edge Nodes, without needing to forward specific logs to a Destination for storage first.

Data buckets are partitioned into subdirectories for Pods and containers, but it will be faster to search for your data by Pods.

Enable disk spooling: Determine whether or not to save Kubernetes logs to disk. When set to **Yes**, the configuration fields below become available.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`). Kubernetes logs are high-volume, so consider the time span you enter here (as a long time span could cause the disk to fill up quickly).

Data size limit: This is the maximum amount of disk space that spooled data is allowed to consume. Once data reaches this amount, Edge will delete data starting with the oldest buckets.

Data age limit: The duration of time for which Cribl Edge will retain data. Edge will delete data that exceeds the max age entered in this field. Example values are `2h`, `4d`. The default value is 24 hours (`24h`).

Compression: The codec that Edge uses to compress the data. The default is `gzip`.

Cribl Edge will write events to the default location:

```
$CRIBL_SPOOL_DIR/in/kube_logs/{inputId}/{timeBucket}/{namespace}/{pod}/{container}
```

By default, By default, `$CRIBL_SPOOL_DIR` points to `$CRIBL_HOME/state/spool`.

To change the base spool path, use the `CRIBL_SPOOL_DIR` environment variable.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

These are the full objects from the Kubernetes API, showing the namespace, node, and Pod where the log originated:

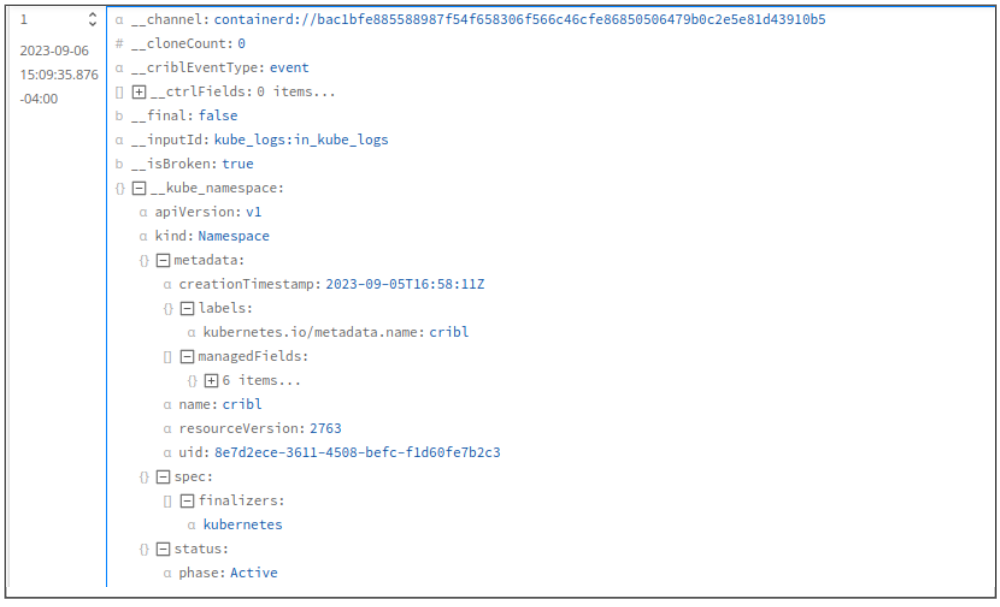
- `__channel`
- `__extractedTime`
- `__final`
- `__isBroken`
- `__kube_namespace`
- `__kube_node`
- `__kube_pod`
- `__raw`
- `__source`
- `_raw`
- `_time`

For details, see Objects in Kubernetes.

Sample Log Event

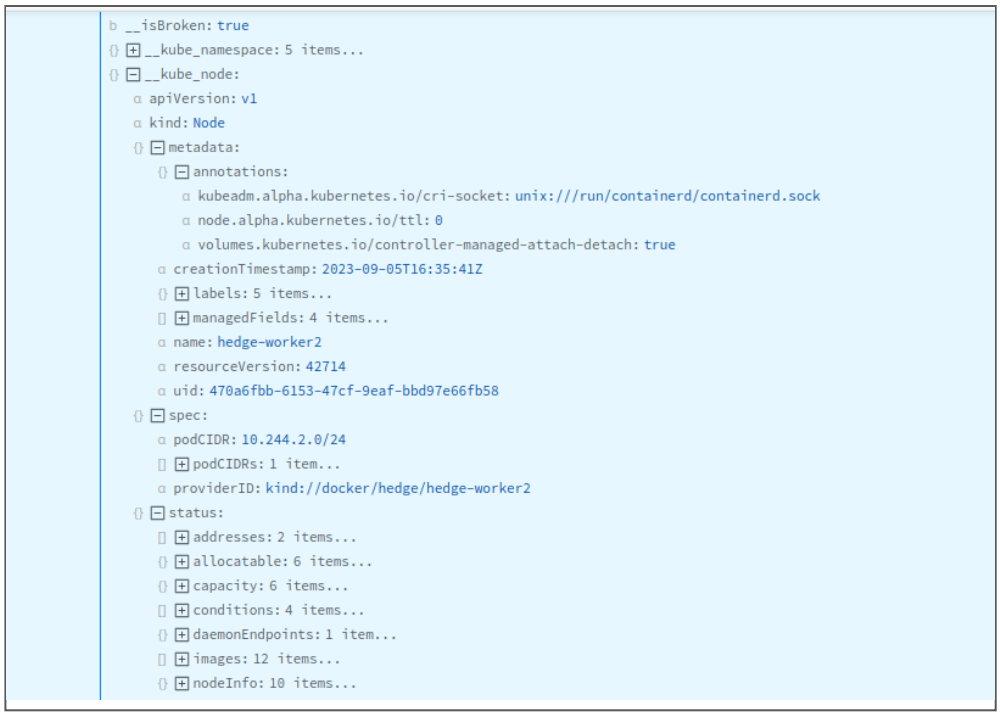
The following screenshots show the `__kube` objects corresponding to a sample log event.

Namespace Metadata



Namespace Metadata

Node Metadata



Node Metadata

Pod Metadata

Pod Metadata

10.1.6. Kubernetes Metrics

The Kubernetes Metrics Source generates events periodically based on the status and configuration of a Kubernetes cluster and its nodes, pods, and containers. For a sample of the events this Source emits, see Live Data.

You must authorize this Source to read metric information from many resources in your cluster. For details, see RBAC for Kubernetes Metrics Source.

Type: **System and Internal** | TLS Support: **N/A** | Event Breaker Support: **No**

Cribl Edge supports configuring only one Kubernetes Metrics Source per Edge Node and per Fleet. This Source is currently unavailable on Cribl-managed Cribl.Cloud Workers.

Discovering and Filtering Kubernetes Objects to Monitor

The Kubernetes Metrics Source connects to the Kubernetes API and loads the lists of K8s nodes, Pods, and containers in the cluster, on a configurable **Polling interval**. The Source then runs the lists through the **Filter Rules** to determine what to report on.

If no rules are configured, or if all of the rules evaluate to `true`, the Source generates metrics for the object. Conversely, if any of the rules evaluates to `false`, the Source skips the object, and does not generate any metrics for it.

Cluster/Daemonset Best Practices

For the Kubernetes Metrics Source to work as designed, Cribl recommends deploying Cribl Edge as a Daemonset. For details, see Deploying via Kubernetes.

This Source can generate redundant events when running inside the cluster and it isn't able to determine which DaemonSet it's in or which Node it's on. This happens when you run Cribl Edge inside the cluster, but in a way that it can't detect the local Pod/node. In these cases, Edge will emit an error when it initializes this Source.

You can bypass the error state by setting the `CRIBL_K8S_FOOTGUN` environment variable to `true`. However, you will assume the risk that the Kubernetes Metrics Source might make excessive calls to

> the cluster API.

Configuring Cribl Edge to Collect Kubernetes Metrics

From the top nav, click **Manage**, then select a **Fleet** to configure. Then, you have two options:

- To configure via the graphical QuickConnect UI, click **Collect**. By default, a **Kubernetes Metrics** tile appears at left. Hover over it and select **Configure** to open a drawer that provides the options below.
- To configure via the Routing UI, click **More** > **Sources**. From the resulting page's tiles or the **Sources** left nav, select **System and Internal** > **Kubernetes Metrics**. Next, click the default `in_kube_metrics` Source to open a modal that provides the options below.

General Settings

Input ID: This is prefilled with the default value `in_kube_metrics`, which cannot be changed via the UI, due to the single-Source restrictions above.

Optional Settings

Polling interval: How often, in seconds, to collect metrics. If not specified, defaults to `15s`.

Filter Rules: Optionally, specify the Kubernetes objects which this Source should parse to generate events. If you specify no restrictive rules here, the Source will emit all events.

- **Filter expression**: JavaScript expression to filter Kubernetes objects. Filters are based on the Kubernetes Pod Object definition, and evaluated from top to bottom. The first expression evaluating to `false` excludes the Pod from collection. The default filter is `!metadata.namespace.startsWith('kube-')`, which ignores Pods in the kube-* namespace.

Other **Filter expressions** examples include:

- Collect metrics from Pods on a specific Node – `spec.nodeName == 'node1'`
- Ignore all DaemonSets – `metadata.ownerReferences[0].kind != 'DaemonSet'`
- Ignore Pods with specific Container names – `spec.containers[0].name != 'edge'`
- **Description**: Optional description of the rule.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

Select a **Pipeline** (or Pack) from the drop-down to process this Source's data. Required to configure this Source via **Data Routes**; optional to configure via **Collect**/**QuickConnect**.

Disk Spooling

Enable disk persistence: Whether to save metrics to disk. Toggle to `Yes` to expose this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `100 MB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Optionally compress the data before sending. Defaults to `gzip` compression. Select `none` to send uncompressed data.

Path location: Path to write metrics to. Default value is `$CRIBL_HOME/state/kube_metrics`.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

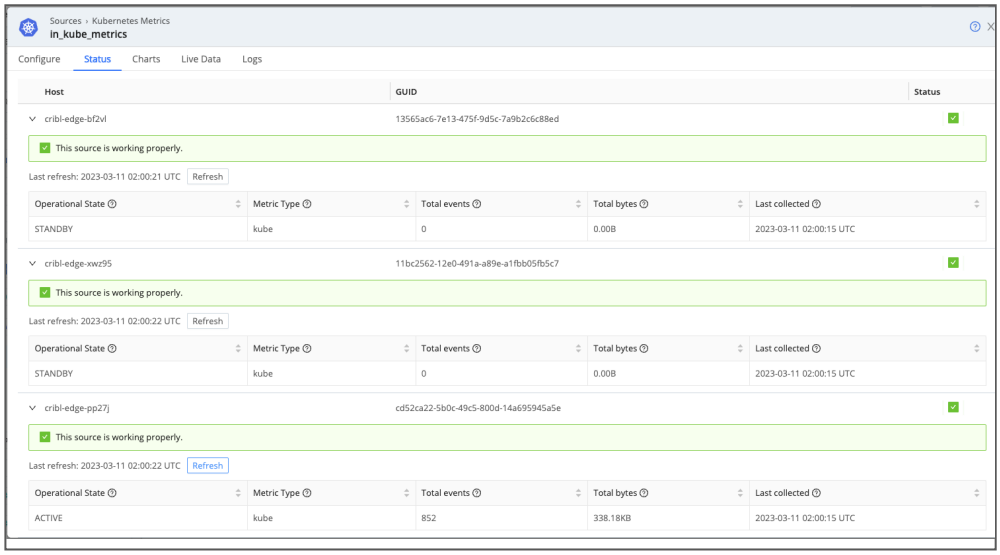
Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Source Operational State

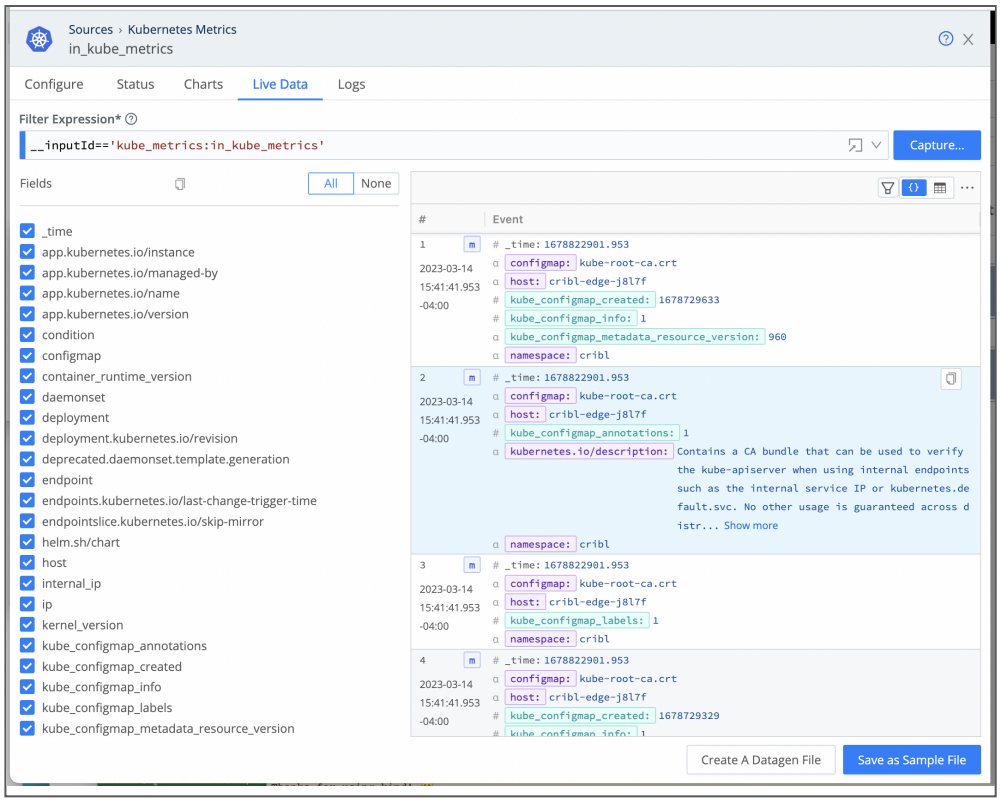
To check the Source's operational state, go to **Status** and expand the host details. The **Operational State** Column shows either an **Active** state or **Standby**. **Active** indicates the Source is running or won the election. **Standby** means it's waiting to be re-elected and not currently running.



Operational State

Source Live Data

To see a sample of the Source's data, click the **Live Data** tab.


Sources › Kubernetes Metrics
in_kube_metrics
Configure
Status
Charts
Live Data
Logs
Filter Expression*
__inputId=='kube_metrics:in_kube_metrics'
Capture...
Fields
All
None
_time
app.kubernetes.io/instance
app.kubernetes.io/managed-by
app.kubernetes.io/name
app.kubernetes.io/version
condition
configmap
container_runtime_version
daemonset
deployment
deployment.kubernetes.io/revision
deprecated.daemonset.template.generation
endpoint
endpoints.kubernetes.io/last-change-trigger-time
endpointslice.kubernetes.io/skip-mirror
helm.sh/chart
host
internal_ip
ip
kernel_version
kube_configmap_annotations
kube_configmap_created
kube_configmap_info
kube_configmap_labels
kube_configmap_metadata_resource_version
#
Event
1
2023-03-14 15:41:41.953 -04:00
_time: 1678822901.953
configmap: kube-root-ca.crt
host: cribl-edge-j8l7f
kube_configmap_created: 1678729633
kube_configmap_info: 1
kube_configmap_metadata_resource_version: 960
namespace: cribl
2
2023-03-14 15:41:41.953 -04:00
_time: 1678822901.953
configmap: kube-root-ca.crt
host: cribl-edge-j8l7f
kube_configmap_annotations: 1
kubernetes.io/description: Contains a CA bundle that can be used to verify the kube-apiserver when using internal endpoints such as the internal service IP or kubernetes.default.svc. No other usage is guaranteed across distr... Show more
namespace: cribl
3
2023-03-14 15:41:41.953 -04:00
_time: 1678822901.953
configmap: kube-root-ca.crt
host: cribl-edge-j8l7f
kube_configmap_labels: 1
namespace: cribl
4
2023-03-14 15:41:41.953 -04:00
_time: 1678822901.953
configmap: kube-root-ca.crt
host: cribl-edge-j8l7f
kube_configmap_created: 1678729329
Create A Datagen File
Save as Sample File

Live Data

10.1.7. System Metrics

Cribl Edge can collect metrics from the host on which it is running, and can populate some standard metrics dashboards right out of the box. To see all of the metrics this Source supports, check out Linux System Metrics Details.

Type: **System** | TLS Support: **N/A** | Event Breaker Support: **No**

Cribl Edge Workers support System Metrics only when running on Linux, not on Windows.

This Source is not available on Cribl-managed Cribl.Cloud Workers.

Configuring Cribl Edge to Collect System Metrics

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

Configure via QuickConnect

1. In the submenu, click **Collect** (Edge only).
2. Click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **System Metrics**.
3. Click either **Add Destination** or (if displayed) **Select Existing**.

The **General Settings** drawer will open.

Configure via Routing

1. Click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge).
2. From the resulting page's tiles or left nav, select [**System and Internal** >] **System Metrics**.
3. Click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Optional Settings

Polling interval: How often to collect metrics, in seconds. Defaults to `10`.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Host Metrics

Use the buttons to select a level of detail.

- **Basic** enables minimal metrics, averaged or aggregated.
- **All** enables full, detailed metrics, specified for individual CPUs, interfaces, and so on.
- **Custom** displays sub-menus and buttons from which you can choose a level of detail (**Basic**, **All**, **Custom**, or **Disabled**) for each type of event.
- **Disabled** means that no metrics will be generated.

The meaning of **All** and **Disabled** are self-evident. **Basic** and **Custom** have different meanings depending on event type, as follows:

System

Basic level captures load averages, uptime, and CPU count.

Custom toggles **Process** metrics on or off; these are metrics for the numbers of processes in various states.

CPU

Basic level captures active, user, system, idle, iowait percentages over all CPUs.

Custom toggles the following on or off: **Per CPU metrics**, **Detailed metrics** (meaning, for all CPU states), and **CPU time metrics** (meaning raw, monotonic CPU time counters).

Memory

Basic level captures captures total, used, available, `swap_free`, and `swap_total`.

Custom toggles **Detailed metrics** on or off. Detailed means for all memory states.

Network

Basic level captures bytes, packets, errors, connections over all interfaces.

Custom exposes the following:

- The **Interface filter**, which specifies which network interfaces to include or exclude (all are included if the filter is empty).
- **Per interface metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off. If on, the **Protocol metrics** toggle appears, allowing you to choose whether to generate metrics for ICMP, ICMPMsg, IP, TCP, UDP, and UDPLite.

Disk

Basic level captures disk used in percent, bytes read and written, and read and write operations, over all mounted disks.

Custom exposes the following:

- The **Device filter**, which specifies which block devices to include or exclude (all are included if the filter is empty). Wildcards and ! (not) operators are supported.
- The **Mountpoint filter**, which specifies which filesystem mountpoints to include or exclude (all are included if the filter is empty). Wildcards and ! (not) operators are supported.
- The **Filesystem type filter**, which specifies which filesystem types to include or exclude (all are included if the filter is empty). Wildcards and ! (not) operators are supported.
- **Per device metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off. If on, the **Enable inode metrics** toggle appears, allowing you to choose whether to generate metrics for filesystem inodes.

Process Metrics

With Process Metrics enabled, Cribl Edge captures process-specific metrics from Linux servers and reports them as events. This allows you to monitor specific processes on Cribl.Cloud instances. You can generate events for any process object.

To collect a process metrics event, create a Process Set and add a filter expression. Processes that match the filter are returned as individual events. See Collecting Metrics. Process Sets are separate from aggregate and host-wide metrics.

Process-specific metrics are **not** affected by the **Host Metrics** detail setting.

Adding a Process Set

To add a Process Set:

1. Open the System Metrics Source and access the **Configure** tab.
2. Click the **Process Metrics** menu, then **Add process set**.
3. Configure the details:
    - **Set name**: The name for this process set.
    - **Filter expression**: The JavaScript expression that will filter the processes.
    - **Include children**: When toggled to "Yes", the processes that match the filter include metrics for child processes.

Filtering Processes

You can filter processes using the field names and values from each process object.

Tip

To see all processes currently running on an Edge Node, click **Explore** in the submenu and open the Processes Tab.

Example filters

This filter looks for Java processes in the sleep state (an interruptible wait) that are using more than 100% of the CPU or more than 25% of memory:

```
cmdline.args[0] == 'java' && status == 'S' && (cpu > 100 || memory_percent > 25)
```

This filter looks for running NGINX web servers that have more than 12 open file descriptors:

```
service == 'nginx' && fds > 12
```

Here are some more Linux processes you could find metrics for:

- `cmdline.args[0].endsWith('/java') && cmdline.args[1] == 'nifi'`
- `exe.path == '/usr/bin/sshd'`
- `uid == 2`
- `cgroup.0 == /user.slice/user-1000.slice/`
- `stat.status == R`
- `stat.starttime > 943517`
- `cpu > 30`
- `memory_percent > 30`

- `stat.num_threads > 10`

The **Filter expression** field expects a specific syntax in order to apply the filter to the processes. These expressions evaluate to either `true` or `false`. See Filters for more information.

Collecting Metrics

Once you have at least one Process Set created and saved, Cribl Edge will begin collecting process metrics at the interval defined in the **Polling interval** field (**General Settings** menu).

To view the status of the Collector and total events collected, click the **Status** tab in the Linux Metrics Source.

To view a live capture of individual metrics gathered from processes that match the Process Set, click the **Live Data** tab. The Process Set that owns each metric is represented by the `__process_set` internal field.

Supported Processes

To view the complete table of supported Linux process metrics, their descriptions, types, and dimensions, see Process-Specific Metrics.

Using Advanced Mode

To test your filter expression against a sample input, click the **Advanced mode** button within the **Filter expression** field. The **Filter expression** modal will open.

Here, you can paste in your filter expression, select a sample JSON input from the drop-down (or enter your own directly in the **Sample input** field), and select a specific event to test against.

The **Advanced mode** window, for process-specific metrics

The **Output** will show you if the filter expression returns any matching processes, based on the filter expressions and JSON input.

Container Metrics

Use the buttons to select a level of detail.

- **Basic** generates the server event and per running container events.
- **All** adds per-device and detailed metrics.
- **Custom** provides controls for customizing which containers to generate metrics from.
- **Disabled** means that no metrics will be generated.

The **Custom** buttons displays additional controls, as follows:

- **Container Filters**: Enter a filter expression to govern which containers to generate metrics from. Leave empty (the default) to generate metrics from all containers.
- **All containers**: Toggle to `Yes` to include stopped and paused containers.
- **Per device metrics**: Toggle to `Yes` to generate separate metrics for each device.
- **Detailed metrics**: Toggle to `Yes` to generate full container metrics.

The **Basic**, **All**, and **Custom** buttons provide the following **Advanced Settings**, which are different from those in the main **Advanced Settings** tab:

- **Docker socket**: Enter the full path(s) for Docker's UNIX domain socket. Defaults to `/var/run/docker.sock` and `/run/docker.sock`.
- **Docker timeout**: Timeout, in seconds, for the Docker API. Defaults to `5`.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

If desired, choose a **Pipeline** from the drop-down if you want to process data from this Source before sending it through the Routes.

Disk Spooling

Enable disk persistence: Whether to save metrics to disk. When set to `Yes`, exposes this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `100 MB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Optionally compress the data before sending. Defaults to `gzip` compression. Select `none` to send uncompressed data.

Path location: Path to write metrics to. Default value is `$CRIBL_HOME/state/system_metrics`.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Populating Dashboards with System Metrics

Cribl has configured a Prometheus dashboard to display Cribl Edge System Metrics, and shared the dashboard in the Grafana library, which is public. This is a relatively simple dashboard suitable for showing aggregate metrics. Try this dashboard if you prefer **Basic** mode for most of your metrics.

Another useful dashboard is the one that's commonly used with Prometheus and their Node Exporter agent, found here. This dashboard can handle the highly detailed metrics. To use this dashboard, attach the `prometheus_metrics` pre-processing pipeline, and choose **All** mode.

For details on how populate your Grafana Cloud dashboards with system metrics, see System Metrics to Grafana.

10.1.8. System State

The System State Source collects snapshots of the host system's current state, on a configurable schedule, and sends them out as events to downstream systems for operational and security analytics.

Type: **System** | TLS Support: **N/A** | Event Breaker Support: **No**

Cribl Edge supports configuring only one System State Source per Edge Node and per Fleet.

This Source is currently unavailable on Cribl-managed Cribl.Cloud Workers.

Configuring Cribl Edge to Collect System State Events

From the top nav, click **Manage**, then select a **Fleet** to configure. Then, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **System State**. By default, a **System State** tile appears at left. Hover over it and select **Configure** to open a drawer that provides the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **System State**. Next, click the default `in_system_state` Source to open modal that provides the options below.

General Settings

Input ID: This is prefilled with the default value `in_system_state`, which cannot be changed via the UI, due to the single-Source restrictions above.

Optional Settings

Polling interval: How often, in seconds, to collect system state events. If not specified, defaults to `300s`.

Each run of the state collector generates events with identical `_time` values.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Collector Settings

Cribl Edge contains the following System State collectors:

- Host Info
- Disks and File Systems
- DNS
- Firewall
- Hosts File
- Interfaces
- Listening Ports
- Logged-In Users
- Routes
- Services
- Users and Groups

Host Info

With the **Enabled** toggle on (the default), Cribl Edge will create events based on entries collected from the hosts file.

A partial list of Host Info events includes:

- `host`
- `timestamp`
- `cribl.version`
- `cribl.mode`
- `cribl.group`
- `cribl.config_version`
- `os.arch`
- `os.cpu_count`

See the **Explore** > **System State** > **Host Info** UI for a full list of events.

Disks and File Systems

With the **Enabled** toggle on (the default), Cribl Edge will collect entries on physical disks, partitions, and mounted filesystems. These entries can help you monitor drive status, including available space, percent utilization, inode availability, wait times, and number of blocks free.

This option is available for Windows and Linux Fleets, but you must create a different Fleet for each OS. Follow this procedure:

1. Create a Fleet for the appropriate OS.
2. Modify the Fleet by disabling or removing the pre-defined Sources that do not apply to the OS. Remove Windows Event Forwarder or Windows Event Logs and Windows Metrics from Linux Fleets, and remove Linux-specific Sources from Windows Fleets.
3. Modify the Fleet mapping to filter by platform – `platform=='win32'` for Windows or `platform=='linux'` for Linux.
4. Deploy Cribl Edge Edge Nodes and verify that the mapping is correct.

If you are using the free version of Cribl Edge, you will only get a single default Fleet. If you need more than one Fleet, you'll have to limit your Cribl Edge deployment to one OS or upgrade to Cribl Edge Enterprise.

For each physical disk, Cribl Edge will collect:

- name
- size (in blocks)
- blockSize (in bytes)
- model
- serial
- firmware
- partitions (count)
- interface type

For each partition on a physical disk, Cribl Edge will collect:

- disk (name)
- disk ID (number)
- type (primary or logical)
- label

- offset
- size (in blocks)
- blockSize (in bytes)
- filesystem (mountpoint)

For each mounted filesystem, Cribl Edge will collect:

- mountpoint (path)
- disk (name)
- partition (ID)
- filesystem type
- size (in blocks)
- blockSize (in bytes)
- blocksFree
- inodes
- inodesFree
- flags

DNS

With the **Enabled** toggle on (the default), Cribl Edge will create events for DNS resolvers and search entries. The event field mapping depends on the operating system.

Linux

Cribl Edge emits events with the following fields:

Event Field	Description
`nameServer`	Name server entries from the file.
`search`	Search entries from the file.
`source`	The source of the data used to populate the event.

Windows

Cribl Edge emits events with the following fields:

Event Field	Description
`ifAlias`	Interface alias name (if present).
`ifIndex`	Interface index (if present).
`addressFamily`	IPv4 or IPv6.
`nameServer`	Array of name servers.
`search`	Array of search suffixes .
`source`	The source of the data used to populate the event.

Hosts File

With the **Enabled** toggle on (the default), Cribl Edge will collect entries from the `hosts` file and emit them as events.

Examples of entries in the `hosts` file are:

- `ip`: IP address of the host.
- `hostnames`: list of hostnames per IP address.

Firewall

With the **Enabled** toggle on (the default), Cribl Edge will collect events from host's defined firewall rules.

Chain Policy/References Rule(s) Event Fields

Cribl Edge emits events with the following fields:

- `chain`: Name.
- `policy`: Policy name.
- `pkts`: Total bytes packets processed by the chain.
- `bytes`: Total bytes processed by the chain.
- `references`: Number of chains referencing the current chain.

Chain Rule(s) Event Fields

Cribl Edge emits events with the following fields:

- `chain`: Name.
- `num`: Rule sequence number.
- `pkts`: Number of matched messages processed.
- `bytes`: Cumulative packet size processed (bytes).
- `target`: If the message matches the rule, the specified target is executed.
- `prot`: The specified protocol can be one of `tcp`, `udp`, `udplite`, `icmp`, `icmpv6`, `esp`, `ah`, `sctp`, `mh`, or the special keyword `all`.
- `opt`: Rarely used, this column is used to display IP options.
- `in`: Inbound network interface.
- `out`: Outbound network interface.
- `src`: The source IP address or subnet of the traffic, the latter being anywhere.
- `dest`: The destination IP address or subnet of the traffic, or anywhere.
- `match`: `ctstate`, `state`, `ADDRTYPE` info added to a state field.
- `comment`: Comment defined on the rule.
- `prot_family`: `IPv4` or `IPv6`.

Windows

On Windows, this collector emits events with the following fields:

- `ruleName`: Rule Name.
- `displayName`: Display Name.
- `direction`: Specifies which direction of traffic to match with this rule. The acceptable values for this parameter are `Inbound` or `Outbound`.
- `enabled`: `True` or `False`.
- `action`: Specifies the action to take on traffic that matches this rule. The acceptable values for this parameter are `Allow` or `Block`.
- `group`: Specifies the rule group (if applicable).
- `icmpType`: Specifies the `ICMP` type used.
- `policyStoreSource`: Contains a path to the policy store where the rule originated.
- `profile`: Profile conditions associated with a rule.
- `protocol`: protocols associated with firewall and `IPsec` rules.
- `localPort`: Rules for local-only port mapping.
- `remotePort`: Rules for remote port mapping.

Interfaces

With the **Enabled** toggle on (the default), Cribl Edge will create events for each of the host's network interfaces.

Cribl Edge will create separate events for each entry. All events will have identical timestamp values.

Linux IPV4/IPV6

On Linux, Cribl Edge creates `interface` events with the following fields:

- `ifIndex`: Interface index identification numbers.
- `ifName`: The name of the network interface.
- `flags`: Flags.
- `mtu`: Maximum transmission unit (MTU) in bytes for packets sent on this interface.
- `operstate`: Operational State.
- `linkType`: Physical link type.
- `macAddress`: Media Access Control address.
- `broadcast` Broadcast address.
- `addrInfo`: Array of `{family, ipAddress, mask, prefix}`.

Windows: IPV4/IPV6

On Windows, Cribl Edge creates `interface` events with the following fields:

- `ifIndex`: Interface index identification numbers.
- `interface`: The name of the network interface.
- `mtu`: Maximum transmission unit (MTU) in bytes for packets sent on this interface.
- `flags`: Flags.
- `macAddress`: Media Access Control address.
- `addrInfo`: Array of `{family, ipAddress, mask, prefix}`.

Listening Ports

With the **Enabled** toggle on (the default), Cribl Edge will create events from the list of listening ports and their associated process identifier (pid).

Cribl Edge will create separate events for each entry. All events will have identical timestamp values.

Sample entries include:

- `Protocol Family`: `ipv4` or `ipv6`.
- `Protocol`: `TCP`, `UDP`, or `UNIX`.
- `Socket`: `${addr}:${port}`, or `${path}` for UNIX sockets.
- `Process id`: The process identifier.
- `Program`: The program listening on the port.

Logged-In Users

With the **Enabled** toggle on (the default), Cribl Edge will collect entries from currently logged-in users.

Linux

Cribl Edge emits events with the following fields:

- `uid`: User ID.
- `Last Login date`: Last login date.
- `terminal`: Type of the terminal device.
- `hostname`: Display the hostname of the system, if the user is connected from a remote computer.
- `userName`: Logon name of the user.

Windows

Cribl Edge emits events with the following fields:

- `userName`: Login name of the user.
- `sessionName`: `rdp` (Remote desktop/terminal services login session) or `console` (Direct login session).
- `id`: Session ID.
- `state`: `active` (Active session) or `disc` (Disconnected/inactive session).
- `logonTime`: Date and time the user logged on.
- `source`: Set to `Query User`.

Routes

With the **Enabled** toggle on (the default), Cribl Edge will collect entries from host's network routes and emit them as events.

Sample entries include:

- `address_family`: `ipv4` or `ipv6`.
- `destination`: The destination network or host.
- `prefix`: Emits only from Linux IPv6.
- `flags`: Flags.
- `gateway`: The gateway address.
- `interface`: Interface to which packets for this route will be sent.
- `ifIndex`: Interface ID. Emits only from Windows.
- `mask`: The netmask for the destination net. Emits only from Linux IPv4.
- `metric`: The distance to the target.
- `sequence`: Sequence number in the table.
- `source`: The source of the data used to populate the event.

Services

With the **Enabled** toggle on (the default), Cribl Edge will create events for each configured service (e.g. `systemd` and `initd`) along with their enabled and running status.

Sample entries include:

- `Name`: Name of the service.
- `Unit Activation Status`: The high-level unit activation state, i.e. generalization of `SUB` for `systemctl`. For `sysV` system value can be `enabled` or `disabled`.
- `Unit Load Status`: Reflects whether the unit definition was properly loaded. (Only `systemctl`).
- `Sub Unit Activation Status`: The low-level unit activation state, values depend on unit type.
- `Description`: Description of the service.

Users and Groups

With the **Enabled** toggle on (the default), Cribl Edge will create events for users and groups.

Linux

On Linux, Cribl Edge creates `user` events with the following fields:

- `username`

- `fullname`
- `loginShell`
- `userHome`
- `userid`
- `primaryGroup` (name of primary group, not group ID)
- `groups` (list of group names)

Cribl Edge also creates `group` events with the following fields:

- `name`
- `groupId`
- `users`

Windows

On Windows, Cribl Edge creates `user` events with the following fields:

- `userName`
- `fullname`
- `description`
- `groups`

Cribl Edge also creates `group` events with the following fields:

- `name`
- `description`
- `users`

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

Select a **Pipeline** (or Pack) from the drop-down to process this Source's data. Required to configure this Source via **Data Routes**; optional to configure via **Collect/QuickConnect**.

Disk Spooling

Enable disk spooling: Whether to save metrics to disk. When set to `Yes`, exposes this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `100 MB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Currently locked to `none`. Cribl plans to add (optional) disk-spooling compression in a future release.

Path location: Path to write metrics to. Default value is `$CRIBL_HOME/state/system_state`.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

10.2. Internal

10.2.1. Cribl HTTP

The Cribl HTTP Source receives data from a Cribl HTTP Destination in the same Distributed deployment, including Cribl.Cloud, at no charge. It's common to send data from Edge and Stream within the same deployment using a Cribl HTTP Destination and Cribl HTTP Source pair.

The Cribl HTTP Source is available only in Distributed deployments. In single-instance mode or for testing, you can substitute it with the Raw HTTP/S Source. However, this substitution will not facilitate sending all internal fields, as described below.

Type: **Internal** | TLS Support: **YES** | Event Breaker Support: **No**

You might choose this Source over the Cribl TCP Source in certain circumstances, such as when a firewall or proxy blocks TCP traffic.

How It Works

You can use the Cribl HTTP Source to transfer data between Workers. If the Cribl HTTP Source receives data from its Cribl HTTP Destination counterpart on another Worker, you're billed for ingress only once – when Cribl first receives the data. All data subsequently relayed to other Workers via a Cribl HTTP Destination/Source pair is not charged.

This use case is common in hybrid Cribl.Cloud deployments, where a customer-managed (on-prem) Node sends data to a Worker in Cribl.Cloud for additional processing and routing to Destinations. However, the Cribl HTTP Destination/Source pair can similarly reduce your metered data ingress in other scenarios, such as on-prem Edge to on-prem Stream.

As one usage example, assume that you want to send data from one Node deployed on-prem, to another that is deployed in Cribl.Cloud. You could do the following:

- Create an on-prem File System Collector (or whatever Collector or Source is suitable) for the data you want to send to Cribl.Cloud.
- Create an on-prem Cribl HTTP Destination.
- Create a Cribl HTTP Source, on the target Stream Worker Group or Edge Fleet in Cribl.Cloud.
- For an on-prem Node configure a File System Collector to send data to the Cribl HTTP Destination, and from there to the Cribl HTTP Source in Cribl.Cloud.

- On Cribl-managed Cribl.Cloud Nodes, make sure that TLS is either disabled on both the Cribl HTTP Destination and the Cribl HTTP Source it's sending data to, or enabled on both. Otherwise, no data will flow. On Cribl.Cloud instances, the Cribl HTTP Source ships with TLS enabled by default.

Configuration Requirements

The key points about configuring this architecture are:

- The Cribl HTTP Destination must be on a Node that is connected to the same Leader as the Cribl HTTP Source(s).
- The Destination's **Cribl endpoint** field must point to the **Address** and **Port** you've configured on its peer Cribl HTTP Source(s).
- Cribl 3.5.4 was a breakpoint in Cribl HTTP Leader/Worker communications. Nodes running the Cribl HTTP Source on Cribl 3.5.4 and later can send data only to Nodes running v.3.5.4 and later. Nodes running the Cribl HTTP Source on Cribl 3.5.3 and earlier can send data only to Nodes running v.3.5.3 and earlier.

Configuring the Cribl HTTP Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select **System and Internal** > **Cribl HTTP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tile or left nav, select [**System and Internal** >] **Cribl HTTP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Cribl HTTP Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the address to bind on. Defaults to `0.0.0.0` (all addresses).

Port: Enter the port number to listen on, e.g., `10200`.

Optional Settings

Tags: Optionally, add tags that you can use for filtering and grouping in the Cribl Stream UI. Use a tab or hard return between (arbitrary) tag names. These tags aren't added to processed events.

TLS Settings (Server Side)

Enabled Defaults to `No`. When toggled to `Yes`, exposes this section's remaining fields.

Certificate name: Select a predefined certificate from the drop-down. A **Create** button is available to create a new certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`, exposes these two additional fields:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common name**: Regex matching subject common names in peer certificates allowed to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. E.g., to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

Enable Persistent Queue defaults to `No`. When toggled to `Yes`:

Mode: Choose a mode from the drop-down:

- With `Smart` mode, PQ will write events to the filesystem only when it detects backpressure from the processing engine.

- With `Always On` mode, PQ will always write events directly to the queue before forwarding them to the processing engine.

Max buffer size: Maximum number of events to hold in-memory before dumping them to disk.

Commit frequency: Number of events to send before committing that Stream has read them.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, Cribl Edge stops queueing and applies the fallback **Queue-full behavior**. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre–Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

The Cribl HTTP Source (and the Cribl TCP Source) treat internal fields differently than other Sources do. That's because of the difference in the way that incoming data originates.

Other Sources ingest data that's not coming from Cribl Edge or Stream, meaning that no Cribl internal fields can be present in that data when it arrives at the Source, and the Source is free to add internal fields without clobbering (overwriting) anything that existed already.

By contrast, the Cribl HTTP Source and the Cribl TCP Source ingest data that's coming from a Cribl HTTP or Cribl TCP Destination. That data **can** contain internal fields when it arrives at the Source. This means that if the Source adds internal fields, those could potentially clobber what existed before.

To avoid this problem, the Cribl HTTP Source and the Cribl TCP Source add a unique `__forwardedAttrs` (i.e., "forwarded attributes") field. The nested structure of the `__forwardedAttrs` field contains any of the following fields that are present in the arriving data:

Internal Fields
`__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
`__inputId`
`__outputId`
`__srcIpPort` – See details below.

Other Fields
`cribl_breaker`
`cribl_pipe`

These fields are **copied** into `__forwardedAttrs`, not moved there. As the data (apart from `__forwardedAttrs`) moves through the Source and any Pipelines, the values of these fields can be overwritten. But the copies of these fields in `__forwardedAttrs` remain unchanged, so you can retrieve them as necessary.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the HTTP client sending data to this Source.

When any proxies (including load balancers) lie between the HTTP client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

10.2.2. Cribl TCP

The Cribl TCP Source receives data from a Cribl TCP Destination in the same Distributed deployment, including Cribl.Cloud, at no charge. It's common to send data from Edge and Stream within the same deployment using a Cribl TCP Destination and Cribl TCP Source pair.

The Cribl TCP Source is available only in Distributed deployments. In single-instance mode or for testing, you can substitute it with the TCP JSON Source. However, this substitution will not facilitate sending all internal fields, as described below.

Type: **Internal** | TLS Support: **YES** | Event Breaker Support: **No**

You might choose this Source over the Cribl HTTP Source in certain circumstances, such as when a firewall or proxy allows TCP traffic.

How It Works

You can use the Cribl TCP Source to transfer data between Workers. If the Cribl TCP Source receives data from its Cribl TCP Destination counterpart on another Worker, you're billed for ingress only once – when Cribl first receives the data. All data subsequently transferred to other Workers via a Cribl TCP Destination/Source pair is not charged.

This use case is common in hybrid Cribl.Cloud deployments, where a customer-managed (on-prem) Node sends data to a Worker in Cribl.Cloud for additional processing and routing to Destinations. However, the Cribl TCP Destination/Source pair can similarly reduce your metered data ingress in other scenarios, such as on-prem Edge to on-prem Stream.

As one usage example, assume that you want to send data from one Node deployed on-prem, to another that is deployed in Cribl.Cloud. You could do the following:

- Create an on-prem File System Collector (or whatever Collector or Source is suitable) for the data you want to send to Cribl.Cloud.
- Create an on-prem Cribl TCP Destination.
- Create a Cribl TCP Source, on the target Stream Worker Group or Edge Fleet in Cribl.Cloud.
- For an on-prem Node, configure a File System Collector to send data to the Cribl TCP Destination, and from there to the Cribl TCP Source in Cribl.Cloud.
- On Cribl-managed Cribl.Cloud Nodes, make sure that TLS is either disabled on both the Cribl TCP Destination and the Cribl TCP Source it's sending data to, or enabled on both. Otherwise, no data will flow. On Cribl.Cloud instances, the Cribl TCP Source ships with TLS enabled by default.

Configuration Requirements

The key points about configuring this architecture are:

- The Cribl TCP Destination must be on a Node that is connected to the same Leader as the Cribl TCP Source.
- When you configure the Cribl TCP Destination, its **Address** and **Port** values must point to the **Address** and **Port** you've configured on the Cribl TCP Source.
- Cribl 3.5.4 was a breakpoint in Cribl TCP Leader/Worker communications. Nodes running the Cribl TCP Source on Cribl 3.5.4 and later can send data only to Nodes running v.3.5.4 and later. Nodes running the Cribl TCP Source on Cribl 3.5.3 and earlier can send data only to Nodes running v.3.5.3 and earlier.

Configuring the Cribl TCP Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Cribl TCP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **Cribl TCP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Cribl TCP Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for TCP JSON data. E.g., `localhost` or `0.0.0.0`.

Port: Enter the port number to listen on, e.g., `10300`.

Additional Settings

Tags: Optionally, add tags that you can use for filtering and grouping in the Cribl Stream UI. Use a tab or hard return between (arbitrary) tag names. These tags aren't added to processed events.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common name**: Regex matching subject common names in peer certificates allowed to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. E.g., to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

Enable Persistent Queue defaults to `No`. When toggled to `Yes`:

Mode: Choose a mode from the drop-down:

- With `Smart` mode, PQ will write events to the filesystem only when it detects backpressure from the processing engine.
- With `Always On` mode, PQ will always write events directly to the queue before forwarding them to the processing engine.

Max buffer size: Maximum number of events to hold in memory before dumping them to disk. Defaults to `1000`.

Commit frequency: Number of events to send before committing that Cribl Edge has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, Cribl Edge stops queueing and applies the fallback **Queue-full behavior**. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre–Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

The Cribl TCP Source (and the Cribl HTTP Source) treat internal fields differently than other Sources do. That's because of the difference in the way that incoming data originates.

Other Sources ingest data that's not coming from Cribl Edge or Stream, meaning that no Cribl internal fields can be present in that data when it arrives at the Source, and the Source is free to add internal fields without clobbering (overwriting) anything that existed already.

By contrast, the Cribl TCP Source and the Cribl HTTP Source ingest data that's coming from a Cribl TCP or Cribl HTTP Destination. That data **can** contain internal fields when it arrives at the Source. This means that if the Source adds internal fields, those could potentially clobber what existed before.

To avoid this problem, the Cribl TCP Source and the Cribl HTTP Source add a unique `__forwardedAttrs` (i.e., "forwarded attributes") field. The nested structure of the `__forwardedAttrs` field contains any of the following fields that are present in the arriving data:

Internal Fields
`__srcIpPort`
`__inputId`
`__outputId`

Other Fields
`cribl_breaker`
`cribl_pipe`

These fields are **copied** into `__forwardedAttrs`, not moved there. As the data (apart from `__forwardedAttrs`) moves through the Source and any Pipelines, the values of these fields can be overwritten. But the copies of these fields in `__forwardedAttrs` remain unchanged, so you can retrieve them as necessary.

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

This kind of periodic logging helps you determine whether a Source is in fact still healthy even when no data is coming in.

10.2.3. Cribl Stream (Deprecated)

This Source was deprecated as of Cribl Edge 3.5, and has been removed as of v.4.0. Please instead use the Cribl TCP or the Cribl HTTP Source to enable Edge Nodes to send data to peer Nodes.

The Cribl Stream internal Source is available only in distributed deployments. It is provided to facilitate sending data between Edge Nodes that are connected to the same Leader. You'll find this Source especially valuable in a hybrid Cloud deployment.

This Source can receive data only from a TCP JSON Destination, on a Edge Node that is connected to the same Leader. The following instructions cover configuring such a TCP JSON Destination.

Type: **Internal** | TLS Support: **YES** | Event Breaker Support: **No**

Use New Sources Instead

The replacement Sources, Cribl TCP and the Cribl HTTP, don't rely on IP filtering, for the following reasons:

- Load balancers and/or proxies between the Cribl Destination and Cribl Source can change the IP address, resulting in a bad match and rejected ingest.
- A Lookup table of all IP addresses needed to be sent to each Worker Node/Edge Node from the Leader, which is not scalable.
- The Lookup table of IP addresses required constant communication between the Worker Node/Edge Nodes and the Leader, making this fragile and placing an arbitrary reliance on the Leader that shouldn't be there.

How It Works

You can use the Cribl Stream Source to send data between Workers. In a hybrid Cribl.Cloud deployment, using this Source ensures that you're billed for ingress only once – when the data is originally received. All other data transferred into Workers via the Cribl Stream Source is not charged.

This use case is common in deployments where a customer-managed (on-prem) Worker sends data to a Worker in Cribl Cloud, for additional processing and then routing to Destinations.

As one usage example, assume that you want to send data from one Worker Node/Edge Node deployed on-prem, to another that is deployed in Cribl Cloud. You could do the following:

- Create an on-prem File System Collector (or whatever Collector or Source is suitable) for the data you want to send to Cribl Cloud.
- Create an on-prem TCP JSON Destination.
- Create a “Cribl Stream” Source, on the target Worker Group/Fleet in Cribl Cloud.
- Configure these to send data from the File System Collector to the TCP JSON Destination, and from there to the “Cribl Stream” Source in Cribl Cloud.

The key points about configuring this architecture are:

- The TCP JSON Destination must be on a Worker Node/Edge Node that is connected to the same Leader as the “Cribl Stream” Source.
- The TCP JSON Destination’s **Address** and **Port** must match the “Cribl Stream” Source’s **Address** and **Port**.

Configuring the Cribl Stream Source

In the **QuickConnect** UI: Click **New Source** or **Add Source**. From the resulting drawer’s tiles, select [**System and Internal** >] **Cribl Stream**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The drawer will now provide the following options and fields.

Or, in the **Data Routes** UI: From the top nav of a Cribl Stream instance or Group, select **Data** > **Sources**. From the resulting page’s tiles or the **Sources** left nav, select [**System and Internal** >] **Cribl Stream**. Next, click **New Source** to open a **Cribl Stream** > **New Source** modal that provides the following options and fields.

General Settings

Input ID: Enter a unique name to identify this Cribl Stream Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for TCP JSON data. E.g., `localhost` or `0.0.0.0`.

Port: Enter the port number to listen on.

Authentication Settings

Use the **Authentication method** buttons to select one of these options:

- **Manual**: Use this default option to enter the shared secret that clients must provide in the `authToken` header field. Exposes an **Auth token** field for this purpose. (If left blank, unauthenticated access will be permitted.) A **Generate** link is available if you need a new secret.

- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the `authToken` header field value described above. The secret can reside in Cribl Stream's internal secrets manager or (if enabled) in an external KMS. A **Create** link is available if you need a new secret.

Optional Settings

Tags: Optionally, add tags that you can use for filtering and grouping in the Cribl Stream UI. Use a tab or hard return between (arbitrary) tag names. These tags aren't added to processed events.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

IP Allowlist Regex: Regex matching IP addresses that are allowed to establish a connection. Although it appears in the **Advanced Settings** UI, this is not a user-configurable setting. Its value is updated automatically to match the IP addresses used in the deployment, and cannot be edited.

Max Active Connections: Maximum number of active connections allowed per Worker Process. Use `0` for unlimited.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Stream uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Field for this Source:

- `__inputId`
- `__srcIpPort`

10.2.4. Cribl Internal

The Cribl Internal Source enables you to capture and send Cribl Edge's own internal **logs** and **metrics** through Routes and Pipelines.

Type: **Internal** | TLS Support: **N/A** | Event Breaker Support: **No**

Scope and Purpose

In distributed mode, this Source's CriblLogs option can process internal logs only from Worker Processes. (Logs on the Leader remain on the Leader, because the Leader Node is not part of any processing path.)

In both distributed and single-instance mode, this Source's CriblLogs option omits API Process logs, meaning that it omits telemetry/license-validation traffic. You can, however, use a Script Collector to check for API Server (or Worker Group) events.

This Source's CriblMetrics option offers Cribl's most up-to-date and precise view of event throughput and transformation, aggregated every 2 seconds at the Worker Process level. The CriblLogs option is less granular – tracking logs that are written once per minute – so CriblLogs might fail to reflect Worker Process crashes or restarts.

In Cribl.Cloud, the CriblLogs internal Source is only available on hybrid, customer-managed Edge Nodes.

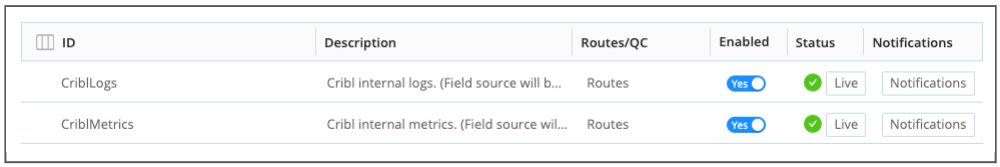
Configuring Cribl Internal Logs/Metrics as a Data Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Cribl Internal**. Next, click **Select Existing**, and click either **Cribl Logs** or **Cribl Metrics** to access the configuration options listed below. When prompted, confirm that you want to switch the existing Source to QuickConnect.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **Cribl Internal**. Next, click either **Cribl Logs** or **Cribl Metrics** to open a modal that provides the configuration options listed below.

In either UI, after you've adjusted or confirmed configuration options: On the **CriblLogs** and/or the **CriblMetrics** row, toggle **Enabled** to `Yes`. Confirm your choice in the resulting message box. The screenshot below shows both options successfully enabled.



Cribl Internal Sources – click either or both of these rows to configure

CriblLogs – General Settings

Enabled: This duplicates the parent page's **Enabled** toggle. Keep it at `Yes` to enable Cribl logs as a Source.

Input ID: Enter a unique name to identify this CriblLogs Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

CriblLogs – Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

CriblMetrics – General Settings

Enabled: This duplicates the parent page's **Enabled** toggle. Keep it at `Yes` to enable Cribl metrics as a Source.

Input ID: Enter a unique name to identify this CriblMetrics Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

CriblMetrics – Optional Settings

Metric name prefix: Enter an optional prefix that will be applied to metrics provided by Cribl Edge. The prefix defaults to `cribl.logstream.`

If Cribl Edge detects `source` or `host` fields in metrics, it copies their values into new dimensions with added `event_` prefixes (e.g., `event_source`). This preserves the original fields' values if they're overwritten in downstream services. For details, see Duplicated Fields/Dimensions.

You can disable metric collection for these and other fields by specifying them in **Settings** > **System** > **General** > **Limits** > **Metrics** > **Disable field metrics**.

Full fidelity: Toggle this to `No` to exclude granular metrics that can cause high CPU load.

The `No` option will drop the following metrics events:

- `cribl.logstream.host.(in_bytes,in_events,out_bytes,out_events)`
- `cribl.logstream.index.(in_bytes,in_events,out_bytes,out_events)`
- `cribl.logstream.source.(in_bytes,in_events,out_bytes,out_events)`
- `cribl.logstream.sourcetype.(in_bytes,in_events,out_bytes,out_events)`

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality. E.g., here you could specify adding an `index` field.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Reporting Metrics Less Frequently

By default, Cribl Edge generates internal metrics every 2 seconds. To consume metrics at longer intervals, you can use or adapt the `cribl_metrics_rollup` Pipeline that ships with Cribl Edge.

Attach this Pipeline to your **Cribl Internal** Source as a pre-processing Pipeline. The Pipeline's **Rollup Metrics** Function has a default **Time Window** of 30 seconds, which you can adjust to a different granularity as needed. This provides a second lever to reduce granularity, in addition to the Full fidelity toggle described above.

Omitting `sourcetype`

You can easily drop the `sourcetype` attribute from metrics events, leaving only `event_sourcetype`. This will prevent duplicate `sourcetype` events from being routed to Destinations.

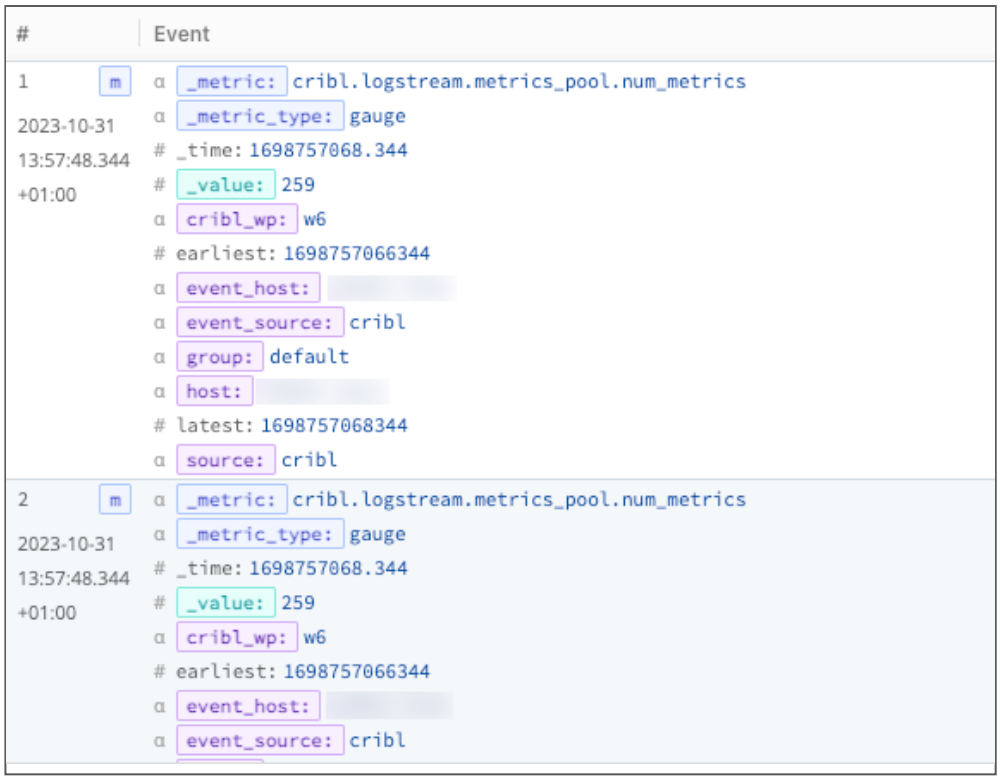
To do this: In the same `cribl_metrics_rollup` pre-processing Pipeline (or a clone) that you attach to your Source, enable the final Eval Function, which applies this **Filter** expression to remove the `sourcetype` field: `_metric && _metric.startsWith('cribl.logstream.sourcetype.')`

Duplicated Fields/Dimensions

The CriblMetrics Source operates on metrics that Edge Nodes report to their Leader Nodes. Typically included are `source` and `host` fields.

Sending metrics from this Source to Splunk is one common use case. Because Splunk might overwrite these two fields, the Source copies their values into new dimensions with added `event_` prefixes: `event_source` and `event_host`. This way, if Splunk does overwrite `source` and/or `host`, their original values remain intact in the new dimensions with `event_` prefixes.

Here's an example of how the added dimensions look in the **Live Capture** window:



Doubled fields

If you are not sending to a downstream service that overwrites `source` or `host` fields, you can use an appropriate Function to drop the added dimensions. (You also have the option to suppress these fields entirely, as covered above in Optional Settings.)

Internal Fields

The following fields will be added to all events/metrics:

- `source`: set to `cribl`.
- `host`: set to the hostname of the Cribl instance.

Use these fields to guide these events/metrics through Cribl Routes.

All Cribl internal fields are subject to change and modification. Cribl provides them to assist with analytics and diagnostics, but does not guarantee that they will remain available.

10.2.5. Datagen

Cribl Edge supports generating data from datagen files, as detailed in Using Datagens. When a datagen is enabled, each Worker Process uses the specified data generator file to generate events. These events proceed through Routes and Pipelines, or through a QuickConnect configuration, to configured Destinations. Whichever Worker Process generated an event from the file will also send the same event.

Type: **System** | TLS Support: **N/A** | Event Breaker Support: **No**

Configuring Cribl Edge to Generate Sample Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Datagen**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **Datagen**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Datagens: List of datagens.

- **Data generator file**: Name of the datagen file.
- **Events per second per Worker Node**: Maximum number of events to generate per second, per Worker Node/Edge Node. Defaults to `10`.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__inputId`

10.3. Amazon

10.3.1. Amazon Kinesis Firehose

Cribl Edge supports receiving data from Amazon Kinesis Data Firehose delivery streams via Kinesis' **HTTP endpoint** destination option.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source supports gzip-compressed inbound data when the `Content-Encoding: gzip` connection header is set.

Configuring Cribl Edge to Receive Data over HTTP(S) from Amazon Kinesis Firehose

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Amazon** > **Firehose**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Amazon** > **Firehose**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Address to bind on. Defaults to 0.0.0.0 (all addresses).

Port: Enter the port number to listen on.

Authentication Settings

Auth tokens: Shared secrets to be provided by any client (Authorization: <token>). Click **Generate** to create a new secret. If empty, unauthenticated access **will be permitted**.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields for this Source:

- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__raw`
- `__srcIpPort` – See details below.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Kinesis Firehose client sending data to this Source.

When any proxies (including load balancers) lie between the Kinesis Firehose client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Limitations/Troubleshooting

If you set the optional `IntervalInSeconds` and/or `SizeInMBs` parameters in the Kinesis Firehose `BufferingHints` API, beware of selecting extreme values (toward the ends of the API's supported ranges). These can send more bytes than Cribl Edge can buffer, causing Cribl Edge to send HTTP 500 error responses to Kinesis Firehose.

10.4. Prometheus

10.4.1. Prometheus Edge Scraper

Cribl Edge supports receiving batched data from Prometheus targets.

Type: **Internal** | TLS Support: **No** | Event Breaker Support: **No**

This Source currently does not support Prometheus metadata.

This is an Internal (Pull) Source. To ingest Prometheus streaming data, see Prometheus Remote Write.

In addition to the functionality already supported in the existing Prometheus Scraper, this Source is designed to work seamlessly in Kubernetes environments; and no longer uses internal jobs framework allowing it to handle large-scale Cribl Edge deployments.

With v.4.2.x, the deprecated Prometheus Scraper is no longer available on Cribl Edge. Please use this Source instead.

Configuring Cribl Edge to Scrape Prometheus Data

From the top nav, click **Manage**, then select a Fleet to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click **Collect**. Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **Prometheus Scraper (Edge)**. Next, click either **Add Source** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **More** > **Sources**. From the resulting page's tiles or left nav, select [**System and Internal** >] **Prometheus Edge Scraper**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Discovery type: Use this drop-down to select a discovery mechanism for targets. See Discovery Type below for the options and the resulting controls displayed.

Some **Discovery type** options replace this section's **Targets** field with additional controls – while also adding an **AWS IAM** left tab to the modal.

Poll interval: Specify how often (in seconds) to scrape targets for metrics. Defaults to `15` seconds. This value must be an integer that divides evenly into `60`.

Discovery Type

Use this drop-down to select a discovery mechanism for targets. To manually enter a targets list, use Static (the default). To enable dynamic discovery of endpoints to scrape, select DNS or AWS EC2. Each selection exposes different controls and/or tabs, listed below.

Static Discovery

The `Static` option adds a **General Settings** > **Targets** field, in which you enter a list of specific Prometheus targets from which to pull metrics. Click **Add Target** to expose a table with the following options:

- **Protocol**: Select `http` (the default) or `https` as the protocol to use when collecting metrics.
- **Host**: Specify the host name to pull metrics from.
- **Port**: Specify the port number to append to the metrics URL for discovered targets. Defaults to `9090`.
- **Path**: Specify a path to use when collecting metrics from discovered targets. Defaults to `/metrics`.

DNS Discovery

The `DNS` option adds the following extra fields to its **General Settings** tab:

- **Record type**: Select the DNS record type to resolve. Defaults to `SRV` (Service). Other options are `A` or `AAAA`.
- **DNS names**: Enter a list of DNS names to resolve.
- **Protocol**: Select `http` (the default) or `https` as the protocol to use when collecting metrics.
- **Path**: Specify a path to use when collecting metrics from discovered targets. Defaults to `/metrics`.

AWS EC2 Discovery

The `AWS EC2` option adds AWS IAM to the modal, and adds extra fields to the **General Settings** tab:

- **Protocol**: Select `http` (the default) or `https` as the protocol to use when collecting metrics.

- **Port**: Specify the port number to append to the metrics URL corresponding to discovered targets. Defaults to `9090`.
- **Path**: Specify a path to use when collecting metrics from discovered targets. Defaults to `/metrics`.
- **Region**: Select the AWS region in which to discover EC2 instances with metrics endpoints to scrape.
- **Use public IP**: The `Yes` default uses the public IP address for discovered targets. Toggle to `No` to use a private IP address.
- **Search filter**: Click **Add filter** to apply filters when searching for EC2 instances. Each filter row provides two columns:
  - **Filter name**: Select standard attributes from the drop-down, or type in custom attributes.
  - **Filter values**: Enter values to match within this row's attribute, Press `Enter` between values. (If you specify no values, the search will return only `running` EC2 instances.)

Selecting `AWS EC2` also adds controls to the **Advanced Settings** tab, as described below.

Kubernetes Node

The `Kubernetes Node` option supports collecting metrics from an endpoint on the local node where Cribl Edge is running. Selecting this option adds the following extra fields to the **General Settings** tab:

- **Protocol**: Select `http` (the default) or `https` as the protocol to use when collecting metrics.
- **Port**: Specify the port number to append to the metrics URL corresponding to discovered targets. Defaults to `9090`.
- **Path**: Specify a path to use when collecting metrics from discovered targets. Defaults to `/metrics`.

Selecting `Kubernetes Node` also adds a control to the **Authentication** tab, as described below.

You must first authorize the Source to discover a Kubernetes Node. For details, see RBAC for Prometheus Edge Scraper.

Kuberenetes Pods

The `Kubernetes Pods` option supports building a list of endpoints to collect from Pods on the same node where Cribl Edge is running. This option adds the following extra fields to the **General Settings** tab – use a custom expression to set the configuration properties on each field.

- **Protocol**: Set the protocol to use when collecting metrics. Defaults to: `metadata.annotations['prometheus.io/scheme'] || 'http'`
- **Port**: Specify the port number to append to the metrics URL corresponding to discovered targets. Defaults to: `metadata.annotations['prometheus.io/port'] || '9090'`
- **Path**: Specify a path to use when collecting metrics from discovered targets. Defaults to: `metadata.annotations['prometheus.io/path'] || '/metrics'`

- **Filter rules**: Optionally, add rules to determine which Pods to discover for metrics. If no rules are provided, Cribl Edge will search all Pods. Otherwise, it will search Pods where rules' expressions evaluate to `true`. Each rule consists of an expression and an optional description.
    - **Filter expression**: JavaScript expression applied to Pods' objects. Return `true` to include it. Filters are based on the Kubernetes Pod Object definition, and are evaluated from top to bottom. The first expression evaluating to `false` excludes a Pod from collection. The default filter is `metadata.annotations['prometheus.io/scrape']`, which scrapes the Pod if the annotation is true.
    - **Description**: Optional description of the rule.

Selecting `Kubernetes Pods` also adds a control to the **Authentication** tab, as described below.

You must first authorize the Source to discover Kubernetes Pods. For details, see RBAC for Prometheus Edge Scraper.

Optional Settings

Extra dimensions: Specify the dimensions to include in events. Defaults to `host` and `source`.

Tags: Optionally, add tags that you can use for filtering and grouping in the Cribl Edge UI. Use a tab or hard return between (arbitrary) tag names. These tags aren't added to processed events.

Authentication (Prometheus)

Use the **Authentication method** drop-down to select one of these authentication options for Prometheus:

- **Manual**: In the resulting **Username** and **Password** fields, enter Basic authentication credentials corresponding to your Prometheus targets.
- **Secret**: This option exposes a **Secret** drop-down, in which you can select a stored secret that references your credentials described above. The secret can reside in Cribl Edge's internal secrets manager or (if enabled) in an external KMS. Click **Create** if you need to configure a new secret.

Authentication Settings for Kubernetes

This additional setting appears only when **General Settings** > **Discovery Type** is set to `Kubernetes Pods` or `Kubernetes Node`.

- **Kubernetes**: Select this option to use Kubernetes authentication. There are no details to configure, because Cribl Edge will authenticate using the Environment Variables or Service Account tokens mounted in the Edge Pod(s). For details, see Deploying via Kubernetes.

AWS IAM

With the AWS EC2 target discovery type, you can configure AssumeRole behavior on AWS.

Assume Role

- **Enable for EC2**: Toggle to `Yes` if you want to use `AssumeRole` credentials to access EC2.
- **AssumeRole ARN**: Enter the Amazon Resource Name (ARN) of the role to assume.
- **External ID**: Enter the External ID to use when assuming the role.

AWS Authentication Options

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Edge Nodes access to authorized AWS resources. Will use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Edge Nodes not in an AWS VPC, e.g., those running a private cloud. This option displays:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. This option displays:

- **Secret key pair**: Use the drop-down to select a secret key pair that you've configured in Cribl Edge's internal secrets manager or (if enabled) an external KMS. Click **Create** if you need to configure a key pair.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Disk Spooling

Enable disk spooling: Whether to save metrics to disk. When set to `Yes`, it exposes this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `1 GB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Defaults to `gzip`.

Cribl Edge will write metrics to the default location: `CRIBL_HOME/state/spool/in/edge_prometheus/${inputId}`. Use the environment variable `CRIBL_SPOOL_DIR`, to change the default path.

Advanced Settings

HTTP Connection Timeout: The amount of time (in milliseconds) to wait for the HTTP connection to establish before it times out. `1-60000` is the allowable range. Enter `0` to disable the timeout.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Advanced Settings for AWS

These additional settings appear only when **General Settings** > **Discovery Type** is set to `AWS EC2`.

Endpoint: Specify an EC2-compatible service endpoint. If empty, this defaults to AWS' Region-specific endpoint.

Signature version: The signature version to use for signing EC2 requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to reject certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`, the restrictive option.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__source`
- `__isBroken`
- `__inputId`
- `__final`
- `__criblMetrics`
- `__channel`
- `__cloneCount`
- `__kube_pod` when Discovery Mode is set to **Kubernetes Node**.
- `__kube_node` and `__kube_pod` when Discovery Mode is set to **Kubernetes Pods**.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

10.4.2. Prometheus Remote Write

Cribl Edge supports receiving metric data from Prometheus instances that are configured to send data via the remote write protocol.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source assumes that incoming data is snappy-compressed.

Configuring Cribl Edge to Receive Metrics from Prometheus Remote Write Sources

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Prometheus** > **Remote Write**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Prometheus** > **Remote Write**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP to listen to. Defaults to `0.0.0.0`.

Port: Enter the port number to listen on..

Remote Write API endpoint: Enter the absolute path on which to listen for Prometheus requests. Defaults to `/write`, which will (in this example) expand as: `http://<your-upstream-URL>:<your-port>/write`.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header, or click **Generate** if you need a new token.
- **Auth token (text secret)**: Provide an HTTP token referenced by a secret. Select a stored text secret in the resulting drop-down, or click **Create** to configure a new secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials. Click **Generate** if you need a new password.
- **Basic (credentials secret)**: Provide username and password credentials referenced by a secret. Select a stored text secret in the resulting **Credentials secret** drop-down, or click **Create** to configure a new secret.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If

the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data.

Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.

- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__criblMetrics`
- `__final`
- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`

- `__srcIpPort` – See details below.
- `_time`
- `_value`

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Prometheus Remote Write client sending data to this Source.

When any proxies (including load balancers) lie between the Prometheus Remote Write client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Detecting Metrics' Types

Because Prometheus remote write requests don't specify metrics' types, Cribl Edge applies the following rules to determine the type as we ingest them:

- If the metric's name ends with `_total`, `_sum`, `_count`, or `_bucket`, the type is set to `counter`.
- Otherwise, the metric's type is set to `gauge`.

This is consistent with the type detection practiced by other services implementing the remote write protocol. See, for example, New Relic's and Elastic's documentation.

Note that Cribl Edge supports the `timer` type in addition to `counter` and `gauge`.

10.4.3. Grafana

Cribl Edge supports receiving metric and log data from Grafana Agent instances via the Prometheus remote write specification. The Grafana Agent uses Prometheus for metrics collection and Grafana Loki for log collection.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source assumes that incoming data is snappy-compressed.

Configuring Cribl Edge to Receive Metrics and Logs from Grafana Agent Sources

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Grafana**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Grafana**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP to listen to. Defaults to `0.0.0.0`.

Port: Enter the port number to listen on.

Optional Settings

Remote Write API endpoint: Absolute path on which to listen for Grafana Agent's remote write requests. Defaults to `/api/prom/push`, which will (in this example) expand as: `http://<your-upstream-URL>:<your-port>/api/prom/push`.

Logs API endpoint: Absolute path on which to listen for Loki logs requests. Defaults to `/loki/api/v1/push`, which will (in this example) expand as: `http://<your-upstream-URL>:<your-port>/loki/api/v1/push`.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

The **Authentication** tab provides separate **Loki** and **Prometheus** sections, enabling you to configure these inputs separately. The two sections provide identical options.

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Enter the bearer token that must be included in the authorization header.
- **Auth token (text secret)**: Provide an HTTP token referenced by a secret. Select a stored text secret in the resulting drop-down, or click **Create** to configure a new secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: Provide username and password credentials referenced by a secret. Select a stored text secret in the resulting **Credentials secret** drop-down, or click **Create** to configure a new secret.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new

connections, clients have an incentive to **maximize** connection reuse.

- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__labels` – For log events only – will contain all the labels found in each event's corresponding Loki stream.

- `__srcIpPort` – See details below.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Grafana client sending data to this Source.

When any proxies (including load balancers) lie between the Grafana client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Detecting Metrics' Types

Because Prometheus remote write requests don't specify metrics' types, Cribl Edge applies the following rules to determine the type as we ingest them:

- If the metric's name ends with `_total`, `_sum`, `_count`, or `_bucket`, the type is set to `counter`.
- Otherwise, the metric's type is set to `gauge`.

This is consistent with the type detection practiced by other services implementing the remote write protocol. See, for example, New Relic's and Elastic's documentation.

Note that Cribl Edge supports the `timer` type in addition to `counter` and `gauge`.

10.4.4. Loki

Cribl Edge supports receiving log data from Grafana Loki via an adaptation of the Protobuf (Protocol Buffers) specification.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source assumes that incoming data is snappy-compressed.

Configuring Cribl Edge to Receive Loki Logs Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Amazon** > **Loki**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Amazon** > **Loki**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP to listen to. Defaults to `0.0.0.0`.

Port: Enter the port number to listen on.

Logs API endpoint: Absolute path on which to listen for Loki logs requests. Defaults to `/loki/api/v1/push`, which will (in this example) expand as: `http://<your-upstream-URL>:<your-port>/loki/api/v1/push`.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication type** drop-down to specify how Loki's Promtail agent will authenticate against Cribl Edge:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.
- **Auth token (text secret)**: Provide an HTTP token referenced by a secret. Select a stored text secret in the resulting drop-down, or click **Create** to configure a new secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: Provide username and password credentials referenced by a secret. Select a stored text secret in the resulting **Credentials secret** drop-down, or click **Create** to configure a new secret.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate server certs**: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If

the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data.

Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

> The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__final`
- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__labels` – Will contain all the labels found in each event's corresponding Loki stream.
- `__srcIpPort` – See details below.
- `_time`

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Loki client sending data to this Source.

When any proxies (including load balancers) lie between the Loki client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

10.5. Splunk

10.5.1. Splunk HEC

Cribl Edge supports receiving data over HTTP/S using the Splunk HEC (HTTP Event Collector).

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **YES**

This Source supports gzip-compressed inbound data when the `Content-Encoding: gzip` connection header is set.

Configuring Cribl Edge to Receive Data over Splunk HEC

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Splunk** > **HEC**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Splunk** > **HEC**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with a Splunk HEC Source preconfigured to listen on Port 8088. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this Splunk HEC Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP on which to listen for HTTP(S) data. Such as: `localhost` or `0.0.0.0`. Supports IPv4 and IPv6 addresses.

Port: Enter the port number.

Splunk HEC endpoint: Absolute path on which to listen for the Splunk HTTP Event Collector API requests. Defaults to `/services/collector`.

This single endpoint supports JSON events via `/event`, health checks via `/health`, raw events via `/raw`, and Splunk S2S events via `/s2s`. See the Splunk REST API endpoints documentation and the examples below. The Source will automatically detect where to forward the request.

Optional Settings

Allowed Indexes: List the values allowed in the HEC event index field. Allows wildcards. Leave blank to skip validation.

Splunk HEC acks: Whether to enable Splunk HEC acknowledgments. Defaults to `No`. See Working with HEC Acks below to learn about context and limitations.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Working with HEC Acks

Cribl Edge sends a `200` or similar HTTP response code when it receives an event.

Some senders also send **ack requests**, which differ from events. This type of request asks Cribl Edge to acknowledge delivery of an array of event IDs. (A sender might **require** HEC acks to be enabled. Cribl does not maintain a comprehensive list of senders that require acks – please refer to your sender's documentation.)

How Cribl Edge responds depends on **Optional Settings** > **Splunk HEC acks**:

- When **Splunk HEC acks** is turned on, Cribl Edge will respond to an ack request with an acknowledgement affirming that Cribl Edge received each of the events whose IDs were listed in the array.
- When **Splunk HEC acks** is turned off, Cribl Edge will respond to an ack request with a `400` error and text saying that `ACK` is disabled.

It makes sense to turn **Splunk HEC acks** on for senders that keep TCP connections open while waiting for an ack, because this behavior can exhaust available file descriptors.

There is a caveat to how HEC acks work in Cribl Edge: The ack response simply cites whatever event IDs appeared in the ack request, regardless of whether or not those events were ever received or processed by Cribl Edge. In that sense, what Cribl Edge sends is a "fake ack:" unlike acknowledgements in some other

systems, the ack in itself is meaningless. It mainly serves to spare a sender from keeping TCP connections open needlessly.

To determine whether Cribl Edge successfully ingested an event, pay no attention to acks. Instead, verify that the sender received a `200` or similar HTTP response from the Splunk HEC source.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (such as, the system's CA). Defaults to `No`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Event Breakers

This section defines event breaking rulesets that will be applied, in order, on the `/raw` endpoint.

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maximum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to determine field's value (can be a constant).

Fields specified on the **Fields** tab will normally override fields of the same name in events. But you can specify that fields in events should override these fields' values.

In particular, where incoming events have no `index` field, this Source adds one with the literal value `default`. You can override this value by using **Add Field** to specify an `index` field, and then setting its **Value** to an expression of the following form: `index == 'default' ? 'myIndex' : index`

Fields here are evaluated and applied **after** any fields specified in the Auth Tokens section.

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Auth Tokens

If empty (the default), the Splunk HEC Source will permit client access without an auth token. To generate and/or configure tokens, click **Add Token**, which exposes the following fields:

Token: Shared secret to be provided by any client (Authorization: <token>). Click **Generate** to create a new secret. If empty, unauthenticated access **will be permitted**.

Enable token: Controls the status of the specified authentication token. When set to `Yes`, the specified auth token is active and can be used for client access. If it's set to `No`, the token is disabled, causing it to be skipped and ignored during initialization. This means that any client attempting to use a disabled token will not be granted access.

Disabling a token can be helpful when you need to temporarily stop receiving data from a particular service or application. Disabling a token prevents the service or application from sending data to your endpoint without having to completely delete the token.

See also Periodic Logging for information on how auth tokens affect product logging.

Description: Optional description for this token.

Fields: Fields to add to events referencing this token. Each field is a **Name/Value** pair.

Fields specified on the **Auth Tokens** tab will normally override fields of the same name in events. But you can specify that fields in events should override these fields' values.

In particular, where incoming events have no `index` field, this Source adds one with the literal value `default`. You can override this value by using **Add Field** to specify an `index` field, and then setting its **Value** to an expression of the following form: `index == 'default' ? 'myIndex' : index`

Fields here are evaluated and applied **before** any fields specified in the Fields section.

Advanced Settings

Use Universal Forwarder time zone (S2S only): Leave the default toggle set to `Yes` to have Event Breakers determine the time zone for events based on Universal Forwarder–provided metadata when the time zone can't be inferred from the raw event.

CORS allowed origins: If you need to enable cross-origin resource sharing with Splunk senders, use this field to specify the HTTP origins to which Cribl Edge should send `Access-Control-Allow-*` headers. You can enter multiple domains and use wildcards. This and its companion **CORS allowed headers** option should seldom be needed – see Working with CORS Headers below. Both options are available in Cribl Edge 4.1.2 and later.

CORS allowed headers: List HTTP headers that Cribl Edge will send in a CORS preflight response to your configured (above) **CORS allowed origins** as `Access-Control-Allow-Headers`. Enter `*` to allow all headers.

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__hecToken`
- `__inputId`
- `__s2sVersion` – value can be either `v3` or `v4`. This field is present only when the Source is receiving S2S-encoded payloads.
- `__srcIpPort` – See details below.
- `__TZ`

Universal Forwarder Time Zone

The `__TZ` field uses the universal forwarder time zone to mitigate cases where incoming events have timestamp strings but no time zone information. Event Breakers use the `__TZ` field to derive time zone information, enabling them to set the `_time` field correctly. See Using the UF Time Zone for additional information.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Splunk HEC client sending data to this Source.

When any proxies (including load balancers) lie between the Splunk HEC client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Working with CORS Headers

The **CORS allowed origins** and **CORS allowed headers** settings are needed **only** when you want JavaScript code running in a web browser to send events to the Splunk HEC Source. In this (uncommon) scenario, the code in question sends a preflight request that includes CORS (Cross-Origin Resource Sharing) headers, and the Splunk HEC Source includes CORS headers in its response.

If the settings (and thus the CORS headers in the response) are correct, the browser will allow the code in question to complete requests to the Splunk HEC Source – that is, allow the code to read the Source's responses. (These settings are optional because for some web applications it's sufficient to send requests **without** reading responses.)

The CORS header dance begins when the web browser sends a request with a header stating its origin. What happens next depends on how you've configured the CORS header settings, as explained in the tables below. (For these examples, we'll assume an origin of `https://mycoolwebapp:3001`.)

CORS allowed origins value	What Splunk HEC Source does	Does Browser then allow code to read response?
`*`	Send `Access-Control-Allow-Origin: *` header.	Yes
Multiple wildcard values	If origin value matches any wildcard, send `Access-Control-Allow-Origin: https://mycoolwebapp:3001` header.	Yes
Nothing (not configured)	Send neither an `Access-Control-Allow-Origin` nor an `Access-Control-Allow-Headers` header.	No

The web browser's request might also include an `Access-Control-Request-Headers` header that lists one or more headers that it wants to use in a subsequent `POST` request (that is, the "actual" as opposed to preflight request). The Splunk HEC Source can then respond with an `Access-Control-Allow-Headers` header **if and only if** it's also sending the `Access-Control-Allow-Origin` header.

CORS allowed headers value	What Splunk HEC Source does	Does Browser then allow code to read response?
`*`	Send `Access-Control-Allow-Headers` header with the same list the client sent.	Yes
Multiple wildcard values	If any headers from the client's list match wildcards, send `Access-Control-Allow-Headers` header listing the matching headers.	Yes
Nothing (not configured)	Do not send an `Access-Control-Allow-`	Probably not

CORS	`Headers` header.	Does Browser then

Format and Endpoint Examples

Splunk HEC to Cribl Edge

- Configure Cribl Edge to listen on port `8088` with an auth token of `myToken42`.
- Send a payload to your Cribl Edge receiver.

Note: Token specification can be either `Splunk <token>` or `<token>`.

Splunk HEC - JSON Event Examples | Splunk HEC - Raw Event Example | Splunk HEC - Health Event Example

```
curl -k http://<myCriblHost>:8088/services/collector/event -H 'Authorization: m

curl -k http://<myCriblHost>:8088/services/collector -H 'Authorization: myToken

# Multiple Events
curl -k http://<myCriblHost>:8088/services/collector -H 'Authorization: myToken

# Metrics Events
curl -k http://<myCriblHost>:8088/services/collector/event -H 'Authorization: m

curl -k http://<myCriblHost>:8088/services/collector/event -H 'Authorization: m

# Send the auth token as a query parameter, with no additional configuration
curl -k "http://<myCriblHost>:8088/services/collector/event?token=mToken42" -d
```

Splunk HEC to Cribl.Cloud

- Navigate to Cribl.Cloud's Splunk HEC Source > **Auth Tokens** tab.
- Copy your token out of the **Token** field.
- From the command line, use `https`, your Cribl.Cloud portal's **Ingest Endpoint** and port, and the token's value:

Splunk HEC > Cribl Cloud endpoint

```
curl -k "https://default.main-<Your-Org-ID>.cribl.cloud:8088/services/collector" \
    -H "Authorization: <token_value>" \
    -d '{"event": "Goats are better than ponies."}{"event": "Goats are better climb
```

With a Cribl.Cloud Enterprise plan, generalize the above URL's `default.main` substring to `<group-name>.main` when sending to other Fleets.

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

If one or more auth tokens are configured and enabled, Cribl Edge logs requests and events for each enabled auth token individually. Since the tokens themselves are redacted for security, Cribl Edge logs the initial text of the token description to help you identify which token a given log is for.

If no auth token is configured and enabled, Cribl Edge simply logs overall statistics about incoming requests and ingested events.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

10.5.2. Splunk TCP

Cribl Edge supports receiving Splunk data from Universal or Heavy Forwarders.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **YES**

Configuring Cribl Edge to Receive Splunk TCP Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Splunk** > **Splunk TCP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Splunk** > **Splunk TCP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with a Splunk TCP Source preconfigured to listen on Port 9997. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this Splunk Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for Splunk data. E.g., `localhost` or `0.0.0.0`.

Port: Enter port number.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

> As of Cribl Edge 4.1, Source-side PQ's default **Mode** changed from `Smart` to `Always on`. This option more reliably ensures events' delivery, and the change does not affect existing Sources' configurations. However:
>
> - If you create Stream Sources programmatically, and you want to enforce the previous `Smart` mode, you'll need to update your existing code.

- If you enable `Always on`, this can reduce data throughput. As a trade-off for data durability, you might need to either accept slower throughput, or provision more machines/faster disks.
- You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Auth Tokens

Add Token : Click to add authorization tokens. Each token's section provides the fields listed below. If no tokens are specified, unauthenticated access **will be permitted**.

Token: Shared secrets to be provided by any Splunk forwarder (Authorization: <token>). Click **Generate** to create a new secret.

Description: Optional description of this token.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (i.e., all IPs).

Max active connections: Maximum number of active connections allowed per Worker Process. Defaults to `1000`. Set a lower value if connection storms are causing the Source to hang. Set `0` for unlimited connections.

Max S2S version: The highest version of the Splunk-to-Splunk protocol to expose during handshake. Defaults to `v3`; `v4` is also available.

Use Universal Forwarder time zone: Displayed (and enabled by default) only when **Max S2S version** is set to `v4`. Provides Event Breakers with a `__TZ` field, which derives events' time zone from UF-provided metadata. See Using the UF Time Zone and Configuring a Splunk Forwarder, below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Using the UF Time Zone

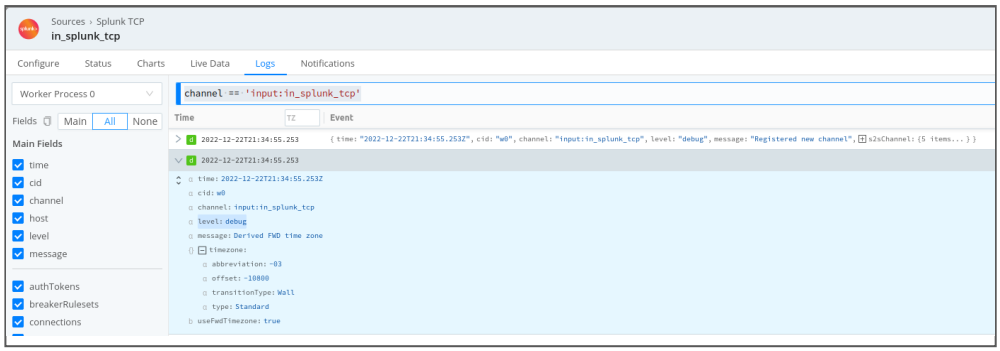
Under Advanced Settings, the **Use Universal Forwarder time zone** toggle mitigates cases where incoming events have timestamp strings but no time zone information. For example:

```
12-15-2022 14:57:22.080 WARN TcpOutputFd [1607 TcpOutEloop] - Connect to 172.17.0.
```

This gap can be problematic, especially if the originating Universal Forwarder is in a different time zone from the processing Edge Node.

The `__TZ` field is the solution. Event Breakers use the `__TZ` field to derive time zone information, enabling them to set the `_time` field correctly. Derived time zone information will appear in Cribl Edge's own logs as

shown below:



Time zone information in logs

Setting the Log Level for Connection Messages

When a Splunk forwarder connects to Cribl Edge, Cribl Edge logs the following message at the `debug` level: `Connection with forwarder has been established successfully.`

To see this message, set the `:forwarders` level to `debug`.

Each message contains details specific to the forwarder, such as the protocol, Splunk version, or remote host, to name a few.

In some situations, logging each incoming connection can produce many messages, which can make it hard to find other messages.

You can adjust the level of these connection messages. To do so, follow these steps:

1. Select **Settings** > **Global Settings** > **Logging** > **Levels**.
2. Search for the channel that logs the connection messages. It will have a name in the form `input:<source-id>:forwarders`. For example: `input:in_splunk_tcp:forwarders`.
3. Set the channel to the log level you prefer, such as `debug` or `silly`.
4. Click **Save** to save your setting.
5. Commit and deploy the change.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__inputId`
- `__s2sVersion` – value can be either `v3` or `v4`
- `__source`
- `__srcIpPort`
- `__TZ` – see above

Configuring a Splunk Forwarder

To configure a Splunk forwarder (UF, HF) use the following sample outputs.conf stanzas:

outputs.conf (on-prem) | outputs.conf (Cribl Cloud)

```
[tcpout]
disabled = false
defaultGroup = cribl, <optional_clone_target_group>, ...
enableOldS2SProtocol = true

[tcpout:cribl]
server = [<cribl_ip>|<cribl_host>]:<port>, [<cribl_ip>|<cribl_host>]:<port>, ..
compressed = false
sendCookedData = true
# As of Splunk 6.5, using forceTimebasedAutoLB is no longer recommended. Ensure
# forceTimebasedAutoLB = false
```

If your use case requires compression, use SSL forwarding to compress the data stream.

With a Cribl.Cloud Enterprise plan, generalize the above URL's `default.main` substring to `<group-name>.main` when sending to other Fleets.

Preventing Data Loss with v3

If you set **Max S2S version** to `v3` and are using Splunk 9.1.0 or later, Cribl recommends that you use the `enableOldS2SProtocol = true` setting shown above to avoid data loss. If you are working with `v3` and a Splunk version earlier than 9.1.0, you should use `negotiateProtocolLevel = 0`. Depending on your environment, enabling `negotiateProtocolLevel` with a non-`0` value could cause Cribl Edge to not accept data from the forwarder.

If you set **Max S2S version** to `v4`, these settings are not necessary. The Splunk receiver will detect which version is in use and automatically use the correct handler.

> See Internal Fields for information on the `__s2sVersion` field.

Troubleshooting Splunk Forwarder Performance Issues

If you encounter performance issues with a Splunk Forwarder, Cribl recommends increasing the number of parallel ingestion Pipelines or increasing forwarder throughput. You can experiment with either or both of these settings.

To increase the number of parallel ingestion Pipelines, adjust the setting for `parallelIngestionPipelines` in `server.conf`. Experiment with values ranging from `2-4`.

To adjust forwarder throughput, increase the `maxKBps` value in `limits.conf`. The default value is `256`. A value of `0` removes all throttling from the forwarder.

10.6. Windows

10.6.1. Windows Event Forwarder

Cribl Edge supports receiving Windows events from the Windows Event Forwarding mechanism built into modern versions of Microsoft Windows (including Windows 10, Windows Server 2012, and more-recent releases).

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source supports client certificates (mutual TLS) authentication and Kerberos authentication on Linux.

On Cribl Edge, this Source currently supports mutual TLS only with Windows Edge Nodes.

Upstream Prerequisites

This page assumes that you:

- Already have Windows Event Forwarding set up.
- Are pointed to one or more Windows Event Collectors (WECs).
- Are using either client certificate authentication over HTTPS or Kerberos authentication over HTTP.

If you are using client certificate authentication, we also assume that you have – or will generate – a server certificate for this Source, issued by the same Certificate Authority as the client certs.

For details, see Configuring Upstream Clients/Senders.

For a complete walk-through of generating certificates, setting permissions and policies, and ingesting Windows Event subscriptions and data through Cribl.Cloud, see our Configuring WEF for Cribl Stream topic.

Configuring Cribl Edge to Receive Windows Events

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Windows Event Forwarder**.

Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Windows Event Forwarder**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Windows Event Forwarder Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP on which to listen for Windows events data. (E.g., `localhost` or `0.0.0.0`.)

Port: Enter the port number. The default, `5986`, is the port used by Windows Event Collector for HTTPS-based subscriptions.

Authentication method: Specify the method for accepting incoming client connections – either `Client certificate` or `Kerebros`.

Client Certification Authentication Settings

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`. If multiple certificates are present in a `.pem`, each must directly certify the one preceding it in that file. This should match the order of traversing the chain from the host to the root CA cert.

The CA certificate that generated the Cribl Edge server certificate must be the same root CA that signed all client certificates that will be forwarding Windows events to this Source. If the issuing CA is an intermediate CA, this CA must also match on the server and all client certs.

Common Name: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If

the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Verify certificate via OCSP: If toggled to `Yes`, Cribl Edge will use an OCSP (Online Certificate Status Protocol) service to check that client certificates presented in incoming requests have not been revoked. Exposes this additional toggle:

Strict validation: If enabled, Cribl Edge will fail checks on **any** OCSP error. Otherwise, Cribl Edge will fail checks only when a certificate is revoked, and will ignore other errors (such as `OCSP server is unavailable` errors).

Kerberos Authentication Settings

To implement Kerberos authentication, you must first prepare your environment.

Service principal name: Specify the full service principal name, in the form `HTTP/<fully qualified domain name>@REALM`. This is case-sensitive.

Keytab location: Specify the path to the keytab file containing the service principal credentials. Cribl Edge will use `/etc/krb5.keytab` if you do not specify a different keytab file.

Kerberos authentication is disabled for Cribl-managed Cribl.Cloud Workers, but it is enabled for on-prem Worker Groups, whether the Leader is based in Cribl.Cloud or on-prem.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Subscriptions

Click **Add Subscription** to define a new subscription. At least one subscription is required, and has the following options:

Name: A friendly name for the subscription, which is only used to help you identify it.

Version: A read-only field which will be populated when the Source is saved, and will be assigned a new value any time new changes are made to the subscription that affect how a client interprets the subscription.

Format: You can choose `Raw` to receive only XML data about the subscribed events, or `RenderedText` to additionally include amplifying information generated by the client about the event's contents.

Heartbeat: The maximum allowable time, in seconds, before the client will check in with Cribl Edge even if it has no new events to send.

Batch timeout: The maximum time, in seconds, that the client should aggregate new events before sending them to Cribl Edge.

Windows Event Collector defines two delivery modes of "Event Delivery Optimization":

- The "Minimize Bandwidth" mode corresponds to a **Heartbeat** and **Batch timeout** of 6 hours (21,600 seconds).
- The "Minimize Latency" mode corresponds to a **Heartbeat** of 1 hour (3,600 seconds) and a **Batch timeout** of 30 seconds.

Read existing events: Control whether historical events should be sent by the client when it first connects. If set to `No`, only events generated by the client after the subscription is received will be forwarded. If set to `Yes`, the behavior depends on the **Use bookmarks** setting:

- If **Use bookmarks** is set to `No`, then each time a client *forcibly* renews its subscription (e.g., after group policy update), all events matching the query will be sent. Restarting Cribl Edge or restarting the client will not cause all events to be re-sent.
- If **Use bookmarks** is set to `Yes`, then when the client receives a subscription for the first time, it will send all historical events matching the query, but on subsequent (even forced) resubscriptions, it will only send events later than the saved bookmark.
- In either case, if a subscription is changed and saved, the saved bookmarks are no longer valid (they are tied to a specific subscription version), and all events matching the updated subscription will be sent.

Use bookmarks: If toggled to `Yes`, Cribl Edge will keep track of which events have been received, resuming from that point even if the client forcibly re-subscribes. If set to `No`, a client (re-)subscribing will either send all historical events, or send only events subsequent to the subscription, depending on the **Read existing events** setting.

Compression: Sets whether Windows clients compress the events sent to Cribl Edge, using the Streaming Lossless Data Compression (SLDC) algorithm. Defaults to `Yes`.

Queries: See the explanation in Configuring Queries below.

Query builder mode: See the explanation in Configuring Queries below.

Targets: Set the DNS names of the clients that should receive this subscription. Wildcard-matching is supported.

Fields: Use this setting to add fields exclusively to the events that WEF delivers **for the subscription you are defining**, using Eval-like functionality. (To add fields to all incoming events regardless of subscription, use **Processing Settings** > **Fields** instead.)

Optionally, you can create the field such that its **Value** is the value of a subscription metadata element. (See the available subscription metadata elements listed below.)

To do this, specify a **Value** of `__subscription.<metadata_element>`. The Source will then find the metadata element you specified in the special object `__subscription` and set the value to that.

For example, if you wanted all events for the subscription to be tagged with its **Locale**, you could set the **Name** to `Locale` (or any arbitrary name) and **Value** to `__subscription.locale`.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Subscription Metadata Elements

Metadata Element	Type	Corresponding UI Setting
`version`	string	**Version**
`subscriptionName`	string	**Name**
`contentFormat`	`Raw` or `RenderedText`	**Format**
`readExistingEvents`	boolean	**Read existing events**
`heartbeatInterval`	number	**Heartbeat**
`batchTimeout`	number	**Batch timeout**
`sendBookmarks`	boolean	**Use bookmarks**
`compress`	boolean	**Compression**

Metadata Element	Type	Corresponding UI Setting
`queries`	array	see About the Query builder mode below

About Query builder mode

This setting produces the `queries` array in one of two different forms, depending on whether you select its **Simple** button or its **Raw XML** button.

Simple produces an array whose elements each contain `path`, a string that corresponds to **Queries** > **Path**, and `queryExpression`, a string that corresponds to **Queries** > **Query expression**.

Raw XML produces an array whose elements each contain only the `xmlQuery` metadata element. This is the XML query you specified, in the form a string, that corresponds to the **XML query** setting.

To access any element in the array, use array indexing. For example:

- `__subscription.queries[0].path` will access the path to the first query in an array created by **Query builder mode** > **Simple**.
- `__subscription.queries[0].xmlQuery` will access the first query in an array created by **Query builder mode** > **Raw XML**.

Configuring Queries

Queries determine which events to send from the clients. **At least one query is required.** The format is derived from the XPath implementation used by Windows Event Collector. You have two **Query builder mode** options: **Simple** or **Raw XML**.

Select **Raw XML** if you have an existing XPath query. Paste your query into the **XML query** field.

Select **Simple** if you need to manually build queries. Click **Add Query** to define each new query. A query has the following properties:

- **Path**: Set this to the `Path` attribute of a `Select` XPath element. See the example below.
- **Query expression**: Set this to the value inside a `Select` XPath element. See the example below.

With either a **Simple** or **Raw XML** query, you can use the **Targets** field to enter the DNS names of the endpoints that should forward these events. Supports wildcards (e.g., `*.mydomain.com`). The default global wildcard (`*`) enables all endpoints.

If you have both **Simple** and **Raw XML** queries defined within a given Subscription, the **Query builder mode** button that you select when saving this Source's configuration determines which

query/queries Cribl Edge will send. To send both **Simple** and **Raw XML** queries, use the **Add Subscription** button to define multiple subscriptions.

Example

When creating a subscription in a Windows Event Collector, you can view the generated XML/XPath query. Consider a subscription that returns all events from the Security log that are of severities Critical, Error, or Warning, and that occurred within the last 24 hours. This subscription would generate XML like this:

```
<QueryList>
  <Query Id="0" Path="Security">
    <Select Path="Security">*[System[(Level=1  or Level=2 or Level=3) and TimeCreat
  </Query>
</QueryList>
```

To use this subscription in a Cribl Edge Windows Event Forwarder Source, you would either use the **Raw XML** option and paste the above XML, or use **Simple** mode and set the following query properties:

- **Path**: `Security`
- **Query expression**: `*[System[(Level=1 or Level=2 or Level=3) and TimeCreated[timediff(@SystemTime) <= 86400000]]]`

You do not need to use the `<Query>` element's `Id` or `Path` attributes anywhere in the Cribl Edge subscription config.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

Use this setting to add fields to **all events coming into the Source**, using Eval-like functionality. (To add fields on a per-subscription basis, use **Subscriptions** > **Fields** instead.)

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Allow MachineID mismatch: Set this to `Yes` if you do not want to verify that the events sent by a client match the client's certificate Common Name (Subject CN). If set to `No`, events where the `MachineID` (by default the client's machine name, like `CLIENT1.domainName.com`) do not match the CN will be rejected.

This setting applies only to client certificate authentication. If you use Kerberos authentication, this option will not appear and will be treated as a `No` value.

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever. See also Adjusting Timeouts.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge how long to wait for additional data before closing the socket connection. Defaults to `90` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the timeout, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

See also Adjusting Timeouts.

CA fingerprint override: Use this setting only if the first certificate in the configured CA chain does not match the SHA1 fingerprint that the WEF client expects. If that is the case, enter the expected SHA1 fingerprint.

This setting only applies to client certificate authentication. It is ignored if you use Kerberos authentication.

Timeout Considerations for Kerberos Authentication

If you use Kerberos authentication, there are some trade-offs you should consider when you select settings in **Subscriptions** > **Batch timeout** and **Advanced Settings** > **Keep-alive timeout (seconds)**.

Windows expects socket connections to be long-lived for Kerberos authentication so it can avoid the overhead of performing the two-step Kerberos authentication on each new batch delivery.

We recommend that you set **Subscriptions** > **Batch timeout**, **Subscriptions** > **Heartbeat**, or both to be less than or equal to **Keep-alive timeout**, especially when a load balancer is in front of the Edge Nodes.

If the socket timeout has elapsed between event deliveries or heartbeats, Cribl Edge will still handle it gracefully, but incur an additional overhead of two network round trips.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__final`
- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__srcIpPort` – See details below.

- `__subscriptionName`
- `__subscriptionVersion`
- `_raw`
- `_time`

Configuring Upstream Clients/Senders

This section summarizes how to configure Windows endpoints/senders to forward events to this Cribl Edge Source. You can find basic instructions for setting up WEF (in a traditional Windows environment) in this Microsoft guide. You can generally follow that guide's "non-domain" section to correctly configure the endpoints/senders.

For a complete walk-through of generating certificates, setting permissions and policies, and ingesting Windows Event subscriptions and data through Cribl.Cloud, see Cribl's Configuring WEF for Cribl Stream topic.

1. Set up Cribl Edge's Windows Event Forwarder Source as outlined above. The CA certificate you use should be the issuing authority both for the server certificate, and for all client certs you plan to have forwarding events to this Source.
2. Ensure that your existing Windows Event Collector is receiving events correctly from whatever endpoints and subscriptions you already have in place.
3. For client certificate authentication only: Find the fingerprint of the CA cert you are using for this Source, via a tool like `certutil` or `certlm` on Windows, or `openssl` on other operating systems. You'll need this in the next step.
4. Edit the group policy for endpoints you want to forward to this Source:
    - Under **Computer Configuration > Administrative Templates > Windows Components > Event Forwarding**, modify the **Configure target Subscription Manager** setting.
    - Add a Subscription Manager with a value like: `Server=https://<cribl-instance>:<wef-source-port>/wsman/SubscriptionManager/WEC,Refresh=<desired refresh interval>,IssuerCA=<CA cert fingerprint from above>`
        - For Kerberos authentication, use `http://`, not `https://`. Kerberos provides its own encryption and does not use TLS.
        - For Kerberos authentication, do not enter the `IssuerCA` portion.
    - Note that the path portion is the same as required for a WEC subscription, and is not configurable in Cribl Edge.
    - Ensure that the protocol is one of the following:

        - `https` for client certificate (mutual TLS) authentication.
        - `http` (for Kerberos authentication).
    - When complete, save the policy and apply it to affected endpoints.
5. Check that events are flowing into Cribl Edge now according to the configured subscriptions. If they are not:
    - Certificate authentication only: Verify that the clients can reach the Cribl Edge instance through the network to port 5986 (or other configured port) via TLS/HTTP. If clients are connecting to Cribl Edge via a proxy, you may need to set **Enable proxy protocol** to `Yes` (in the **Advanced** section of the Source configuration). Ensure that the correct outbound firewall port is opened on the client.
    - Certificate authentication only: Verify all of the following: that the certificate chain is correct; that the endpoints have a valid CRL encompassing the CA cert; that the CA cert is a trusted root on the clients; and that the server and client certs are issued by the same CA. The `CAPI2` Windows event log might reveal any errors here.
    - Certificate authentication or Kerberos authentication: Check Cribl Edge for any errors, as well as the `EventForwarding-Plugin` and `Windows Remote Management` event logs on the clients.

Preparing the Environment for Kerberos Authentication

This section documents the setup steps you need to take before configuring a Windows Event Forwarder Source for Kerberos authentication.

Initial Conditions and Assumptions

We assume that:

- The clients that will be sending events are in a Windows domain. In this example, the domain is `contoso.com`, the NETBIOS name is `CONTOSO`, and the domain controller is `DC01.contoso.com`.
- Kerberos KDC is already configured and working on the domain, and WEC with Kerberos authentication already works on the domain.
- You have an x86-64 Edge Node to which you have shell access. This guide assumes a recent Ubuntu distribution.
- The fully qualified domain name (FQDN) of this example Edge Node is `worker1.contoso.com`.
- You are using a Kerberos credential cache of the `DIR` or `FILE` type.

Initial Setup for the Edge Node

To set up the Edge Node:

1. Ensure that the Edge Node can resolve the domain controller's FQDN. You can test it with: `ping DC01.contoso.com`.
2. Ensure that the Edge Node's `hostname` is set to the FQDN. If it's not, you can set it with `hostnamectl set-hostname worker1.contoso.com`.
3. Ensure the Edge Node has its clock synchronized with the domain controller.
4. Install the `krb5-user` package and configure Kerberos to correctly identify your domain. Here's an example configuration (`/etc/krb5.conf`).

```
[libdefaults]
default_realm = CONTOSO.COM

[realms]
  CONTOSO.COM = {
    kdc = DC01.contoso.com
    admin_server = DC01.contoso.com
  }

[domain_realm]
  .contoso.com = CONTOSO.COM
  contoso.com = CONTOSO.COM
```

Setup for the Domain Controller

To set up the domain controller:

1. Set up forward and reverse DNS for the Edge Node's FQDN (i.e., both A and PTR records).
2. Both `nslookup worker1.contoso.com` and `nslookup <worker's IP>` should work.
3. Create a domain user for the Edge Node. This guide assumes the user is called `svc-stream-worker1`.
4. Set the user's password to never expire.
5. Under the user's **Properties** > **Account** tab, set both `This account supports Kerberos AES 128 bit encryption` and `This account supports Kerberos AES 256 bit encryption`.

Keytab Export

To export the keytab:

1. Export a keytab file for this user, which will also create the appropriate service principal name (SPN). The keytab will be exported to the root of the user profile running the command.
2. Open an elevated PowerShell prompt on the domain controller.
3. Run `ktpass /princ http/worker1.contoso.com@CONTOSO.COM /pass <password> /mapuser CONTOSO\svc-stream-worker1 /crypto AES256-SHA1 /ptype KRB5_NT_PRINCIPAL /out worker1.contoso.com.keytab`.
4. Note the SPN created here (in this example, `http/worker1.contoso.com@CONTOSO.COM`) as you will need it later. This SPN is case-sensitive.
5. Copy the keytab file to the Cribl Edge Edge Node. Carefully note the file's location on the Edge Node, because you will need it later. The keytab command and output will resemble this example:

```
PS C:\Users\Administrator> ktpass /princ http/wefkerb.weftest.local@WEFTEST.LOCAL /
Targeting domain controller: SUP-DC01.weftest.local
Using legacy password setting method

Successfully mapped http/wefkerb.weftest.local@WEFTEST.LOCAL to cribl.
Key created.

Output keytab to wefkerb.weftest.local.keytab:
Keytab version: 0x502
keysize 110 http/wefkerb.weftest.local@WEFTEST.LOCAL ptype 1 (KRB5_NT_PRINCIPAL) vr
```

If you rotate passwords on this user service account, you'll need to export a new keytab file and merge or replace the keytab on each Edge Node. Microsoft describes a general method for this task. The Microsoft method is for a different service, but it explains the general principles. You could also use a tool like `msktutil`.

Setup for the Windows Event Forwarder Source

The general setup for creating subscriptions is the same as documented in Subscriptions.

When you create the Source, select the `Kerberos` authentication option. Specify the SPN and keytab file location that you set above in Keytab Export. The SPN is case-sensitive, and must be exactly in the form used to export the keytab.

Setup for the Windows Event Forwarding Target

To set up the Windows event forwarding target, you'll need to create group policy objects (GPOs) for the subscription target, the WinRM service, and the NETWORK SERVICE.

Create a GPO for the Subscription Target

1. Access the UI **Computer Configuration** > **Administrative Templates** > **Windows Components** > **Event Forwarding** > **Configure target Subscription Manager**.
2. Set to `Enabled`.
3. Go to **Subscription Managers** > **Show**.
4. Add `Server=http://worker1.contoso.com:5985/wsman/SubscriptionManager/WEC,Refresh=60` to the list.

Create a GPO for the WinRM Service

1. If you haven't already done so, create a GPO to auto-start the WinRM service on the client machines:
2. Access the UI **Computer Configuration** > **Policies** > **Windows Settings** > **Security Settings** > **System Services** > **Windows Remote Management (WS-Management)**.
3. Set **Startup type** to `Automatic`.

Create a GPO for the NETWORK SERVICE

1. If you haven't already done so, create a GPO so that the NETWORK SERVICE can read event logs (required if you want to subscribe to the Security event log):
2. Access the UI **Computer Configuration** > **Policies** > **Windows Settings** > **Security Settings** > **Restricted Groups**.
3. In the right pane, right-click and select **Add Group**.
4. Enter `Event Log Readers` as the group name.
5. Under `Members of this group`, click **Add**.
6. Enter `NETWORK SERVICE` as the member name.

For the change to take effect, you'll have to restart computers after you apply this GPO.

When you have finished up your environment, select the appropriate settings for Kerberos authentication.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the WEF client sending data to this Source.

In the WEF Source configuration, there's an option to handle the `X-Forwarded-For` header, which is commonly used by load balancers to pass the original client's IP address. When any proxies (including load balancers) lie between the WEF client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Adjusting Timeouts

For best results, adjust the following timeouts in **Advanced Settings**:

- **Socket timeout** must be higher than **Subscriptions** > **Batch timeout** (or the lowest **Batch timeout** if several subscriptions are used).
- **Keep-alive timeout** must be higher than **Subscriptions** > **Batch timeout** (or the lowest **Batch timeout** if several subscriptions are used).
- **Keep-alive timeout** must be higher than **Subscriptions** > **Heartbeat**.

Troubleshooting

See the troubleshooting section of the Configuring WEF for Cribl Stream topic.

10.6.2. Windows Event Logs

Cribl Edge supports collecting local Windows Event Logs in batches. You can collect the standard event logs – Application, Security, and System – and any other event logs on the machine.

For guidance on workflow, see our better practices doc: Windows Observability Using Cribl Edge.

Type: **Pull** | TLS Support: **N/A** | Event Breaker Support: **No**

Cribl Edge Workers support Windows Event Logs only when running on Windows, not on Linux.

Configuring Cribl Edge to Collect Windows Event Logs

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

Configure via QuickConnect

1. To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge).
2. Click **Add Source** at left. From the resulting drawer's tiles, select **Push** > **Windows Event Logs**.
3. Click either **Add Destination** or (if displayed) **Select Existing**.

The **General Settings** drawer will open.

Configure via Routing

1. Click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge).
2. From the resulting page's tiles or left nav, select **Push** > **Windows Event Logs**.
3. Click **Add Source** to open the **New Source** modal.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Event logs: Enter one or more event logs to collect. `Security` is prefilled as a default. Click **Add log** to specify more event logs (e.g., `Application` or `System`).

- This Source can collect events from the Forwarded Events log when **Event format** is set to **XML**.
- For more information about how to locate Windows event logs on a Windows server, see Locate Windows Events Logs in the Server.

Optional Settings

Read mode: Select **Entire Log** (the default) to read all of the historical events and new events. Select **From last entry** to read only new events.

Event format: Select JSON or XML as the event format. JSON is the default. You can change this setting at any time.

- It may be faster to render events as XML.
- JSON format includes the rendered message string, while XML does not.
  Sample JSON


```
{
  "_raw": "
  {\"Id\":1234,\"Version\":0,\"Qualifiers\":null,\"Level\":0,\"Task\":9999,\"Opcode
  "source": "Security",
  "host": "EC2-HOST",
  "_time": "1000000000.000",
  "cribl_breaker": "windows event logs (json)"
}
```


Sample XML


```
 {
  "_raw": "<Event xmlns='[http://schemas.microsoft.com/win/2004/08/events/event](ht
  "source": "Security",
  "host": "EC2-Host",
  "_time": "1000000000.000",
  "cribl_breaker": "windows event logs (xml)"
 }
```


Polling interval: Specify how often, in seconds, to check for new event logs. Defaults to `10` seconds, and must be at least `1` second. Polling will read each log, up to the **Batch Size**, in their specified order. This Source will not make a polling request if it's still reading events from the previous request.

Batch size: Set the maximum number of events to read per polling request. By default, each request reads up to `500` events. Set as high as you need to; however, if you configure the Source to pull from multiple event logs, be aware that setting **Batch Size** higher can keep one log waiting longer while the Source collects a batch from another log.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Max event bytes: The maximum number of bytes that can be in an event before it is flushed to the pipelines. Defaults to `51200`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Send to Routes: Enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

QuickConnect: Send this Source's data to one or more Destinations via independent, direct connections.

This Source defaults to QuickConnect.

Troubleshooting

If you are running this Source to collect events on a Windows Node without admin access, change the permission on the registry entry for the `EventLog` key: `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\EventLog` to `Read`.

10.6.3. Windows Metrics

Cribl Edge collects metrics from Windows hosts on which it runs, and can populate some standard metrics dashboards right out of the box. This Source produces events compatible with the Prometheus Windows Exporter to send out system, CPU, memory, network, and disk metrics. For metrics details, see Windows System Metrics Details.

Type: **System and Internal** | TLS Support: **N/A** | Event Breaker Support: **No**

Edge Nodes support Windows Metrics only when running on Windows. Edge Leaders support configuring only one Windows Metrics Source per Fleet.

Configuring Cribl Edge/Windows to Collect System Metrics

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

Configure via QuickConnect

1. In the submenu, click **Collect**.
2. Under **Sources**, find the **Windows Metrics** tile. Hover over it and select **Configure**.

The **General Settings** drawer will open.

Configure via Routing

1. Click **More** > **Sources**.
2. From the resulting page's tiles or the **Sources** left nav, select **System and Internal** > **Windows Metrics**.
3. Click the default **Windows Metrics** Source.

The **General Settings** drawer will open.

General Settings

Input ID: This is prefilled with the default value `in_windows_metrics`, which cannot be changed via the UI, due to the single-Source restriction above.

Optional Settings

Polling interval: How often, in seconds, to collect metrics. Defaults to `10` seconds.

Tags: Optionally, add tags that you can use for filtering and grouping in the Cribl Edge UI. Use a tab or hard return between (arbitrary) tag names. These tags aren't added to processed events.

Host Metrics

Use the buttons to select a level of detail:

- **Basic** enables minimal metrics, averaged or aggregated. This is the minimum configuration needed to support the Cribl Edge landing page.
- **All** enables full, detailed metrics for individual CPUs, interfaces, etc.
- **Custom** displays submenus and buttons from which you can choose a level of detail (**Basic**, **All**, **Custom**, or **Disabled**) for each specific type of event.
- **Disabled** generates no metrics.

The meanings of **All** and **Disabled** are self-evident. **Basic** and **Custom** have different meanings depending on event type – see the following subsections.

System

Basic level captures load averages, uptime, and CPU count.

Custom level toggles **Detailed** metrics on or off. These are Windows-specific metrics including OS information, system uptime, CPU architecture, etc.

CPU

Basic level captures active, user, system, idle, and iowait percentages over all CPUs.

Custom level toggles the following on or off: **Per CPU metrics**, **Detailed metrics** (i.e., metrics for all CPU states), and **CPU time metrics** (i.e., raw, monotonic CPU time counters).

Memory

Basic level captures captures total, used, available, `swap_free`, and `swap_total`.

Custom level toggles **Detailed metrics** on or off. (These are metrics for all memory states.)

Network

Basic level captures bytes, packets, errors, and connections over all interfaces.

Custom level exposes the following:

- The **Interface filter**, which specifies which network interfaces to include or exclude. (An empty filter will include all metrics.)
- **Per interface metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off. If on, the **Protocol metrics** toggle appears, allowing you to choose whether to generate metrics for ICMP, ICMPMsg, IP, TCP, UDP, and UDPLite.

Disk

Basic level captures disk usage (%), bytes read and written, and read and write operations, over all mounted disks.

Custom level exposes the following:

- The **Volume filter**, specifying which Windows volumes to include or exclude. Supports wildcards and ! (not) operators. An empty filter will include all volumes.
- **Per volume metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off.

Process Metrics

With Process Metrics enabled, Cribl Edge captures process-specific metrics from Windows servers and reports them as events. This allows you to monitor specific processes on Cribl.Cloud instances. You can generate events for any process object.

To collect a process metrics event, create a Process Set and add a filter expression. Processes that match the filter are returned as individual events. See Collecting Metrics. Process Sets are separate from aggregate and host-wide metrics.

Process-specific metrics are **not** affected by the **Host Metrics** detail setting.

Adding a Process Set

To add a Process Set:

1. Open the Windows Metrics Source and access the **Configure** tab.
2. Click the **Process Metrics** menu, then **Add process set**.
3. Configure the details:
    - **Set name**: The name for this process set.
    - **Filter expression**: The JavaScript expression that will filter the processes.

- **Include children**: When toggled to "Yes", the processes that match the filter include metrics for child processes.

Filtering Processes

You can filter processes using the field names and values from each process object.

Tip

To see all processes currently running on an Edge Node, click **Explore** in the submenu and open the Processes Tab.

Example Filters

Here's an example filter and the results it could return.

This filter will retrieve processes with the name `explorer`:

```
processName === 'explorer'
```

Note that the strict equality (`===`) operator in the above expression forces case sensitivity in the filter.

The **Filter expression** field expects a specific syntax in order to apply the filter to the processes. These expressions evaluate to either `true` or `false`. See Filters for more information.

Collecting Metrics

Once you have at least one Process Set created and saved, Cribl Edge will begin collecting process metrics at the interval defined in the **Polling interval** field (**General Settings** menu).

To view the status of the Collector and total events collected, click the **Status** tab in the Windows Metrics Source.

To view a live capture of individual metrics gathered from processes that match the Process Set, click the **Live Data** tab. The Process Set that owns each metric is represented by the `__process_set` internal field.

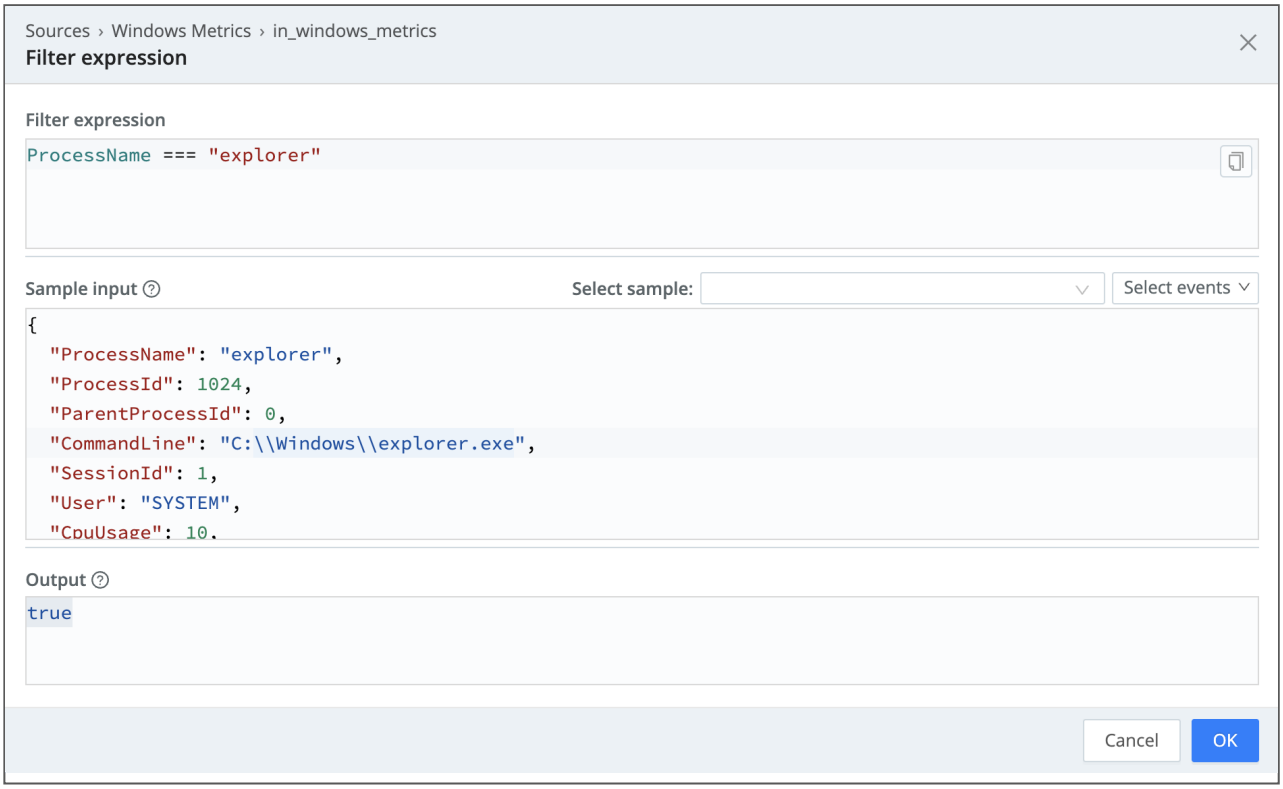
Supported Processes

To view the complete table of supported Windows process metrics, their descriptions, types, and dimensions, see Process-Specific Metrics.

Using Advanced Mode

To test your filter expression against a sample input, click the **Advanced mode** button within the **Filter expression** field. The **Filter expression** modal will open.

Here, you can paste in your filter expression, select a sample JSON input from the drop-down (or enter your own directly in the **Sample input** field), and select a specific event to test against.



The **Advanced mode** window, for process-specific metrics

The **Output** will show you if the filter expression returns any matching processes, based on the filter expressions and JSON input.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

Select a **Pipeline** (or Pack) from the drop-down to process this Source's data. Required to configure this Source via **Data Routes**; optional to configure via **Collect**/**QuickConnect**.

Disk Spooling

Enable disk persistence: Whether to save metrics to disk. When set to `Yes`, exposes this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `100 MB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Optionally compress the data before sending. Defaults to `gzip` compression. Select `none` to send uncompressed data.

Path location: Path to write metrics to. Default value is `$CRIBL_HOME/state/windows_metrics`.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

10.7. AppScope

AppScope is an open-source instrumentation utility from Cribl. It offers visibility into any Linux command or application, regardless of runtime, with no code modification. For details about configuring the AppScope CLI, loader, and library, see: https://appscope.dev/docs. Note that AppScope is no longer being actively developed by Cribl.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **YES**

Configuring Cribl Edge to Receive AppScope Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**System and Internal** >] **AppScope**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**System and Internal** >] **AppScope**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Downloading AppScope

Cribl Edge must download the AppScope package from the Cribl CDN the first time it performs an operation on this Source. This will also update the AppScope package's version, if a newer one is available in the CDN.

If Cribl Edge cannot access the CDN (for example, if you are working on an airgapped instance), you should manually download the binary and place it in `$CRIBL_HOME/state/download/scope`.

General Settings

Input ID: Enter a unique name to identify this AppScope Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.s

UNIX domain socket: When toggled to `Yes`, exposes the following two fields to specify a file-backed UNIX domain socket connection to listen on.

- **UNIX socket path**: Path to the UNIX domain socket. Defaults to `$CRIBL_HOME/state/appscope.sock`.
- **UNIX socket permissions**: Permissions to set for this socket, e.g., `777`. If empty, Cribl Edge will use the runtime user's default permissions.

When **UNIX domain socket** is set to `No`, you instead see the following two fields to specify a network host and port.

- **Address**: Enter the hostname/IP on which to listen for AppScope data. (E.g., `localhost`.) Defaults to `0.0.0.0`, meaning all addresses.
- **Port**: Enter the port number to listen on.

By default:

- In Cribl Stream, **UNIX domain socket** is set to `No`, with default network connections (address and port) of `0.0.0.0:10090` for TCP, and `0.0.0.0:10091` for TLS, respectively.
- In Cribl Edge, **UNIX domain socket** is set to `Yes`, with a UNIX socket path of `$CRIBL_HOME/state/appscope.sock`.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Use this default option to enter the shared secret clients must provide in the `authToken` header field. Click **Generate** if you need a new auth token. If empty, unauthenticated access will be permitted.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the auth token described above. The secret can reside in Cribl Edge's internal secrets manager or (if enabled) in an external KMS. Click **Create** if you need to configure a new secret.

Optional Settings

UNIX socket permissions: Permissions to set for socket. For the preconfigured `in_appscope` source, defaults to `777`. When creating a new AppScope Source, you should set this to `777`. If empty, falls back to the runtime user's default permissions.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

This left tab is displayed only when the Optional Settings > **UNIX domain socket** toggle is set to `No`. It provides the following options.

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

AppScope Rules

The AppScope Rules settings are available:

- In Cribl Stream single-instance – but not Cribl.Cloud – deployments.
- In Cribl Edge, both single-instance and Cloud.

Rules: Click **Add Rule** to include processes to scope, and to link to an AppScope config. Once you have saved the configuration, and committed and deployed your changes, AppScope will instrument any process that matches a Rule, on any Edge Node in the Fleet.

(When no Rules are defined, you can still scope by PID in the Edge Processes page. Scope by PID only instruments a single process running in one Edge Node.)

- **Process name**: Matches if the string value you enter corresponds to the basename of the scoped process.
- **Process argument**: Matches if the string value you enter appears as a substring anywhere in the scoped process' full command line (including options and arguments).
- **AppScope config**: Select an AppScope config.

Transport override: Enter a URL to override aspects of the transport configuration, such as the hostname, port, or TLS settings. For details, see Transport Override Details.

Transport Override Details

In scenarios like the following, use the **Transport override** option to extend the defaults in AppScope's transport configuration:

- When this Source is set to **TCP** mode, it typically listens on the default address `0.0.0.0`. Scoped clients will need a specific `IP` or `hostname` to connect. In these cases, set **Transport override** URL to a specific IP/hostname (example format: `tcp://my.host.name`). This Source will parse the URL and look for the `hostname` and `port`, then use those values to override what would otherwise be sent to the `scope start` command.
- When this Source is set to **UNIX domain socket**, it listens by default on `$CRIBL_HOME/state/appscope.sock`. The socket is often created on a mounted volume in a container. The path to that socket might be different outside the container, or when mounted into another container. In these cases, set the **Transport override** URL to specify an alternative path (example format: `unix:///some/other/volume/appscope.sock`).

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Disk Spooling

For Cribl Search to access the data that arrives at an AppScope Source, **Disk Spooling** must be enabled.

Enable disk persistence: Whether to save metrics to disk. When set to `Yes`, exposes this section's remaining fields.

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`).

Max data size: Maximum disk space the persistent metrics can consume. Once reached, Cribl Edge will delete older data. Example values: `420 MB`, `4 GB`. Default value: `100 MB`.

Max data age: How long to retain data. Once reached, Cribl Edge will delete older data. Example values: `2h`, `4d`. Default value: `24h` (24 hours).

Compression: Optionally compress the data before sending. Defaults to `gzip` compression. Select `none` to send uncompressed data.

Path location: Path to write metrics to. Default value is `$CRIBL_HOME/state/appscope.sock`.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection.

Max active connections: Maximum number of active connections allowed per Worker Process. Defaults to `1000`; enter `0` to allow unlimited connections.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

AppScope with Edge on Kubernetes

When Cribl Edge detects that a `scope`'d process is running inside a Kubernetes container, it reports the Kubernetes metadata as `kube_**` properties. It adds these to the incoming events and metrics, and you can view the combined events and metrics on the AppScope Source's **Live Data** tab.

For Cribl Edge to detect that it's running in a Kubernetes Pod, you must first set the `CRIBL_K8S_POD` environment variable.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__inputId`
- `__srcIpPort`

Examples

The following examples work only in Cribl Stream. You can vary the scoped commands (`ps -ef` and `curl`) as desired.

Cribl.Cloud – TLS

An `in_appscope_tls` TLS Source is preconfigured for you on Cribl.Cloud, using port `10090`. You can send it AppScope data using this command:

```
./scope run -c <Your-Ingress-Address>:10090 -- ps -ef
```

Cribl Cloud – TCP

An `in_appscope_tcp` TCP Source is preconfigured for you on Cribl.Cloud, using port `10091`. You can send it AppScope data using this command:

```
./scope run -c tcp://<Your-Ingress-Address>:10091 -- curl -so /dev/null \
https://wttr.in/94105
```

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

This kind of periodic logging helps you determine whether a Source is in fact still healthy even when no data is coming in.

10.8. Datadog Agent

Datadog Agent is open-source software that monitors the host on which it runs. Acting as a DogStatsD server, Datadog Agent also aggregates metrics from other processes or containers on the host.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

Cribl Edge can ingest the following from Datadog Agent, in bundled form:

- Logs.
- Metrics (gauge, rate, counter, and histogram).
- Service checks.
- Agent metadata and other events emitted from the `/intake/` endpoint.

Datadog Agent also emits Application Performance Monitoring data, which it sends to Datadog. Cribl Edge does not currently support ingesting this APM data.

On the system(s) that you want to monitor, you'll need to install Datadog Agent and configure it to send data to Cribl Edge. Cribl Edge can parse, filter, and enrich that data, and then send it to any supported Destination, including a Cribl Edge Datadog Destination. (By default, Datadog Agent sends data only to Datadog.)

For inbound log data, this Source supports gzip -compression when the `Content-Encoding: gzip` connection header is set. For other data types, it assumes that all inbound data is compressed using `deflate`.

Configure Datadog to use Datadog API v1

This Source communicates with the Datadog API. Although v2 is the latest version of this API, this Source uses v1, which Datadog still supports.

To configure your Datadog Agent to work with this Source, ensure that in the `datadog.yaml` file, `use_v2_api.series` is set to `false`. Otherwise, when Datadog Agent sends metrics, Cribl Edge will not receive them, and `API Key invalid, dropping transaction` errors will appear in the Datadog Agent logs.

Potential Cribl Edge support for Datadog API v2 is being tracked as **CRIBL-18281** in Known Issues.

Configuring Cribl Edge to Ingest Datadog Agent Output

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Datadog Agent**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Datadog Agent**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Datadog Agent Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for Datadog Agent data. E.g., `localhost` or `0.0.0.0`.

Port: Enter the port number to listen on.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests per worker process. Use `0` for unlimited.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

Extract metrics: Toggle to `Yes` to extract each incoming metric to multiple events, one per data point. This works well when sending metrics to a `statsd`-type output. If sending metrics to DatadogHQ or any destination that accepts arbitrary JSON, leave toggled to `No` (the default).

Forward API key validation requests: Toggle to `Yes` to send key validation requests from Datadog Agent to the Datadog API. If toggled to `No` (the default), Stream handles key validation requests by always responding that the key is valid.

Reject unauthorized certificates: Reject certificates that are not authorized by a trusted CA (e.g., the system's CA). Defaults to `No`. Available when **Forward API key validation requests** is toggled to `Yes`.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__agent_api_key`
- `__agent_event_type`

- `__final`
- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__srcIpPort` – See details below.
- `_time`

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Datadog Agent sending data to this Source.

When any proxies (including load balancers) lie between the Datadog Agent and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Sending Datadog Agent Data to Cribl Edge

Before you begin this section, you should have Datadog Agent running on one or more hosts.

To enable Datadog Agent to send data to Cribl Edge, you'll set the following environment variables:

- `DD_DD_URL`: The URL or IP address and port of your Datadog Agent Source in Cribl Edge.
- `DD_LOGS_CONFIG_LOGS_DD_URL`: The same value as above, assuming log collection is enabled on Datadog Agent.
- `DD_LOGS_CONFIG_USE_HTTP`: Set to `true`. Cribl Edge ingests Datadog Agent logs over HTTP only; ingesting logs over TCP is not supported.

Set these environment variables in one of two ways: (1) through a `docker run` command, or (2) by editing the `datadog.yaml` configuration file that your Datadog Agent uses.

Using the `docker run` Command

Here's an example `docker run` that sets the environment variables described above, along with others required for Datadog Agent but not relevant to Cribl Edge. Replace the example values with values

appropriate for your environment.

```
docker run --rm --name dd-agent
  -e DD_API_KEY=6d1ephx1w55p978i6d1ephx1w55p978i \
  -e DD_SKIP_SSL_VALIDATION=true \
  -e DD_DD_URL="http://0.0.0.0:8080" \
  -e DD_LOGS_ENABLED=true \
  -e DD_LOGS_CONFIG_LOGS_DD_URL="0.0.0.0:8080" \
  -e DD_LOGS_CONFIG_USE_HTTP=true \
  -e DD_LOGS_CONFIG_LOGS_NO_SSL=true \
  -e DD_LOGS_CONFIG_CONTAINER_COLLECT_ALL=true \
  -e DD_DOGSTATSD_NON_LOCAL_TRAFFIC="true" \
  gcr.io/datadoghq/agent:7
```

Using a Config File

Instead of using environment variables, you can set the Cribl Edge-related values in your Datadog Agent's `datadog.yaml` config file.

See the documentation links in the example config file that Datadog maintains online; and, the docs that describe filesystem locations relevant to getting Datadog Agent to configure itself with the desired `datadog.yaml` file.

Managing What Data Goes Where

The different kinds of information that Datadog Agents emit - logs, metrics, service checks, agent metadata, and APM data - are not completely independent when you send them to Cribl Edge Datadog Agent Source. The reference to "bundling" near the top of this page introduced this situation in simplified form. The table below explains the relevant constraints for each type of information that Datadog Agents emit, along with the Datadog environment variables that control them.

Type(s) of Information	Dependencies and/or Limitations	Environment Variable
Logs	Independent of other types.	`DD_LOGS_CONFIG_DD_URL`
Metrics, Events, Service Checks, and Metadata	Always sent together.	`DD_DD_URL`
APM Data	Can only be sent to Datadog, e.g., `https://trace.agent.datadoghq.com`.	`DD_APM_DD_URL`

Type(s) of Information	Dependencies and/or Limitations	Environment Variable
	Cannot be sent to Cribl Edge even when you **are** sending other types there.	

Managing API Keys

Suppose your data flow runs from Datadog Agents, to Cribl Edge Datadog Agent Sources, to Cribl Edge Datadog Destinations, to Datadog accounts. You'll need to decide **how many** of each of these elements there are to define the data flow you want. You will also set (or override) **Datadog API keys** to support the desired data flow. For some data flows, you'll need the **General Settings** > **Allow API key from events** toggle in the Cribl Edge Datadog Destination.

Another possibility is that you want data to flow to some Destination other than a Cribl Edge Datadog Destination. If that's the case, try adapting the examples below to your use case. Please connect with us on the `Cribl Community` Slack if you have questions.

Data Flow Examples

The examples which follow start simple and become more complex in terms of desired data flow, and, consequently, of Cribl Edge configuration.

One Agent to One Account

- You're running one Datadog Agent on one host.
- You want the output of the Agent sent to a single Datadog account.
- You only need one API key.
- You configure that single API key on your Cribl Edge Datadog Destination.

Many Agents to One Account

- You're running Datadog Agents on a fleet of hosts.
- You want to consolidate all the output of the Agents into a single Datadog account.
- You only need one API key. This is no different from the simpler one-to-one scenario above.
- You configure that single API key on your Cribl Edge Datadog Destination.
- No matter which host the data originates from, your Cribl Edge Datadog Destination sends it to Datadog using that single API key.

Many Agents to Many Accounts

- You're running Datadog Agents on a fleet of hosts.
- You want the output of the Agents from some hosts to flow into one Datadog account, and the output from other hosts to flow into a different Datadog account.
    - This may need to scale up to multiple accounts, if, for example, you are managing data from several different customers, or from several different organizations within your company.
- You need one API key **for each** Datadog account.
- In the Cribl Edge Datadog Destination, toggle **General Settings** > **Allow API key from events** to `Yes`.
- Here's what happens:
    - Datadog Agent always sends an API key when it communicates with the Cribl Edge Datadog Agent Source. Agents within different subgroups of hosts will send different API keys, as described above.
    - Cribl Edge Datadog Agent Source passes the API key from the Agent to the Cribl Edge Datadog Destination as an internal field.
    - Cribl Edge Datadog Destination uses that API key to direct its output to the correct Datadog account. If need be, you can configure the Destination to override the passed-in API key with a different one. This comes in handy when you know that your Agent(s) are using invalid API keys. You can even override **all** passed-in API keys.

Verifying that Data is Flowing

Once you have configured your Datadog Agent(s) as described above, and they have begun sending data:

- Try viewing the incoming data in the **Live Data** tab in your Cribl Edge Datadog Agent Source.
- If you have a Cribl Edge Datadog Destination set up, connect your Cribl Edge Datadog Agent Source to it via a **QuickConnect Passthru**, and verify that the same data shows up in the Destination's **Live Data** tab.
    - Another possibility is that you want your Cribl Edge Datadog Agent Source to connect to some Destination other than a Cribl Edge Datadog Destination. That's beyond the scope of this example, but verifying data flow should be similar.
- If the Cribl Edge Datadog Destination is configured to send data to a Datadog instance, verify that the same data shows up there, in the appropriate form, according to the transformations you've configured in Cribl Edge.
- If you have a many-to-many data flow as described above, verify that output is being correctly split among the possible Datadog accounts at the end of the data flow.

Assuming that data is behaving as expected, you can proceed to the best part, namely adding one or more Pipelines between Datadog Agent Source and Datadog Destination, to transform the data however you wish. Given the varied nature of what Datadog Agents collect, there should be ample opportunities to make the data leaner and more usable.

10.9. Elasticsearch API

Cribl Edge supports receiving data over HTTP/S using the Elasticsearch Bulk API.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source supports gzip-compressed inbound data when the `Content-Encoding: gzip` connection header is set.

For examples of receiving data from popular senders via this API, see Configuring Elastic Beats below.

Configuring Cribl Edge to Receive Elasticsearch Bulk API Data over HTTP(S)

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Elasticsearch API**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Elasticsearch API**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with an Elasticsearch API Source preconfigured to listen on Port 9200. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this Elasticsearch Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP on which to listen for Elasticsearch data. (E.g., `localhost` or `0.0.0.0`.)

Port: Enter the port number.

Elasticsearch API endpoint (for Bulk API): Absolute path on which to listen for Elasticsearch API requests. Defaults to `/`. Cribl Edge automatically appends `_bulk`, so (e.g.) `/myPath` becomes `/myPath/_bulk`.

Requests could then be made to either `/myPath/_bulk` or `/myPath/<myIndexName>/_bulk`. Other entries are faked as success.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

In the **Authentication type** drop-down, select one of the following options:

- **None**: Don't use authentication.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials. Click **Generate** if you need a new password.
- **Basic (credentials secret)**: Provide username and password credentials referenced by a secret. Select a stored text secret in the resulting **Credentials secret** drop-down, or click **Create** to configure a new secret.
- **Auth tokens**: Use HTTP token authentication. Click **Add Token** and, in the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header. Click **Generate** if you need a new token. Click **Add Token** to display additional rows to specify more tokens.

See also Periodic Logging for information on how auth tokens affect product logging.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Enable proxy mode: If you toggle this to `Yes`, see Proxy Mode below for the resulting options.

Extra HTTP Headers: Name-value pairs to pass as additional HTTP headers. By default, Cribl Edge's responses to HTTP requests include the `X-elastic-product` header, with an `Elasticsearch` value. (This is required by certain clients, including some Elastic Beats.)

API Version: To upstream Elastic Beats, this Cribl Edge Source will appear as an Elasticsearch instance matching the version that you set in this drop-down:

- `6.8.4` – Retained for backward compatibility.
- `8.3.2` – This default entry matches Elasticsearch's current `8.3.x` versions.

- `Custom` – Opens an HTTP response object in the **Custom API Version** editor. This object replicates what an Elasticsearch server would send in its HTTP responses to a client. You can edit the `version : number`, and any other fields, as required to satisfy the Elastic Beat sending data to this Source. (This `Custom` option supports future Elasticsearch releases, as long as Elasticsearch keeps the same response-object structure.)

Proxy Mode

Enabling proxy mode allows Cribl Edge to proxy non–Bulk API requests to a downstream Elasticsearch server. This can be useful when integrating with Elasticsearch API senders like Elastic Endgame agents, which send requests that Cribl Edge does not natively support. Sliding **Enable proxy mode** to `Yes` exposes the following controls.

Proxy URL: URL of the Elasticsearch server that will proxy non-bulk requests, e.g.: `http://elastic:9200`.

Reject unauthorized certificates: Reject certificates that are not authorized by a trusted CA (e.g., the system's CA). Defaults to `No`.

Remove headers: Enter any headers that you want removed from proxied requests. Press `Tab` or `Return` to separate header names.

Proxy request timeout: How long, in seconds, to wait for a proxy request to complete before aborting it. Defaults to `60` seconds; minimum timeout is `1` second.

To understand how **Proxy request timeout** interacts with the `X-Forwarded-For` header, see Overriding `__srcIpPort` with Client IP/Port.

Authentication method: Select one of the following options.

- **None**: Don't use authentication.
- **Manual**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Secret**: This option exposes a **Credentials secret** drop-down, in which you can select a stored secret that references the credentials described above. A **Create** link is available to store a new, reusable secret.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and

tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Field Normalization

The Elasticsearch API input normalizes the following fields:

- `@timestamp` becomes `_time` at millisecond resolution.
- `host` is set to `host.name`.
- Original object `host` is stored in `__host`.

The Elasticsearch Destination does the reverse, and it also recognizes the presence of `__host`.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__host`
- `__id`
- `__index`
- `__inputId`
- `__pipeline` - If present in the Elasticsearch `event.action` field of the event. See also how to override the pipeline attribute.
- `__srcIpPort` – See details below.
- `__type`

Overriding the Pipeline Attribute in the Elasticsearch and Elastic Cloud Destinations

The Elasticsearch Source will record the pipeline attribute received in a variable named `__pipeline`, if pipeline was presented in the Source event. If you want to forward the pipeline from the Source event to an Elasticsearch or Elastic Cloud Destination, set up the **Elastic pipeline** in one of the following ways.

This expression will use the value of `__pipeline` or default to `'myPipeline'` if `__pipeline` is missing:

```
__pipeline || 'myPipeline'
```

An alternative is to pass the pipeline if present, or otherwise omit it:

```
__pipeline ? __pipeline : undefined
```

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the Elasticsearch client sending data to this Source.

When any proxies (including load balancers) lie between the Elasticsearch client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Configuring Elastic Beats

Beats are open-source data shippers that act as agents, sending data to Elasticsearch (or to other services, in this case Cribl Edge). The Beats most popular with Cribl users are Filebeat and Winlogbeat.

To set up a Beat to send data to Cribl Edge, edit the Beat's YAML configuration file: `filebeat.yml` for Filebeat, `winlogbeat.yml` for Winlogbeat, and so on. In the config file, you'll specify your Cribl Edge Elasticsearch Source endpoint as the Beat's Elasticsearch output. To the Beat, Cribl Edge will appear as an instance of Elasticsearch.

If you're using HTTP token authentication (which is disabled by default, both on-prem and on Cribl.Cloud): First, set the token. Then add the following to the Beat config file under `output.elasticsearch.headers`, substituting your token for `myToken42`:

```
output.elasticsearch:
  headers:
    Authorization: "myToken42"
```

Configuring an Elastic Agent

An Elastic Agent is single agent for logs, metrics, security data, and threat prevention that sends data to Elasticsearch (or to other services, in this case Cribl Edge).

When you are sending from an Elastic Agent to Cribl Edge, the `elastic-agent.yml` file doesn't pay any attention to the settings for `output.elasticsarch.allow_older_versions: true`. As a result, the

Elasticsearch API Source will not get any data.

To set it up correctly, go to the **Advanced Settings** tab, and change the **API Version** to `Custom`. In the **Custom API Version** editor, edit the `version : number` to match the version of the Elastic Agent you are using. This should allow the data to start flowing to the Source.

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

If one or more auth tokens are configured and enabled, Cribl Edge logs requests and events for each enabled auth token individually. Since the tokens themselves are redacted for security, Cribl Edge logs the initial text of the token description to help you identify which token a given log is for.

If no auth token is configured and enabled, Cribl Edge simply logs overall statistics about incoming requests and ingested events.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

10.10. HTTP/S (Bulk API)

Cribl Edge supports receiving data over HTTP/S from Cribl Bulk API, Splunk HEC, and Elastic Bulk API endpoints.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

This Source supports gzip-compressed inbound data when the `Content-Encoding: gzip` connection header is set.

Configuring Cribl Edge to Receive Data over HTTP(S)

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **HTTP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **HTTP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with an HTTP Source preconfigured to listen on Port 10080, and on several default endpoints. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this HTTP(S) Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP on which to listen for HTTP(S) data. (E.g., `localhost` or `0.0.0.0`.)

Port: Enter the port number.

Authentication Settings

Auth tokens: Shared secrets to be provided by any client (`Authorization: <token>`). Click **Generate** to create a new secret. If empty, unauthenticated access **will be permitted**.

See also Periodic Logging for information on how auth tokens affect product logging.

Optional Settings

Cribl HTTP event API: Base path on which to listen for Cribl HTTP API requests. To construct the actual endpoint, Cribl Edge will append `/_bulk` to this path. For example, with the default value of `/cribl`, your senders should send events to a `/cribl/_bulk` path. Use an empty string to disable.

Elastic API endpoint (for Bulk API): Base path on which to listen for Elasticsearch API requests. Currently, the only supported option is the default `/elastic`, to which Cribl Edge will append `/_bulk`. So, your senders should send events to an `/elastic/_bulk` path. Other entries are faked as success. Use an empty string to disable.

Cribl generally recommends that you use the dedicated Elasticsearch API Source instead of this endpoint. The Elastic API implementation here is provided for backward compatibility, and for users who want to ingest multiple inputs on one HTTP/S port.

Splunk HEC endpoint: Absolute path on which to listen for Splunk HTTP Event Collector (HEC) API requests. Use an empty string to disable. Default entry is `/services/collector`.

This Splunk HEC implementation is an **event** (i.e., not **raw**) endpoint. For details, see Splunk's documentation. To send data to it from a HEC client, use either `/services/collector` or `/services/collector/event`. (See the examples below.)

Cribl generally recommends that you use the dedicated Splunk HEC Source instead of this endpoint. The Splunk HEC implementation here is provided for backward compatibility, and for users who want to ingest multiple inputs on one HTTP/S port.

Splunk HEC Acks: Whether to enable Splunk HEC acknowledgements. Defaults to `No`.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable proxy protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__headers` – Added only when **Advanced Settings** > **Capture request headers** is set to `Yes`.
- `__inputId`
- `__srcIpPort` – See details below.
- `__host` (Elastic In)
- `__id` (Elastic In)
- `__index` (Elastic In)
- `__type` (Elastic In)

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the client sending data to this Source.

When any proxies (including load balancers) lie between the HTTP client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original HTTP client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

Format and Endpoint

Cribl Edge expects HTTP(S) events to be formatted as one JSON record per event. Here are two event records:

Sample Event Format

```
{"_time":1541280341, "_raw":"this is a sample event ", "host":"myHost", "source":"r
{"_time":1541280341, "host":"myOtherHost", "source":"myOtherSource", "_raw": "{\"me
```

Note 1: Events can be sent as separate POSTs, but Cribl **highly** recommends combining multiple events in newline-delimited groups, and POSTing them together.

Note 2: If an HTTP(S) source is routed to a Splunk destination, fields within the JSON payload are mapped to Splunk fields. Fields that do not have corresponding (native) Splunk fields become index-time fields. For example, let's assume we have a HTTP(S) event like this:

```
{"_time":1541280341, "host":"myHost", "source":"mySource", "_raw":"this is a sample event ", "fieldA":"valueA"}
```

Here, `_time`, `host` and `source` become their corresponding fields in Splunk. The value of `_raw` becomes the actual body of the event, and `fieldA` becomes an index-time field. (`fieldA::valueA`).

Examples

Cribl Edge

The examples in this section demonstrate sending HTTP data into a Cribl Edge binary that you manage on-prem, or on a VM. To set up these examples:

1. Configure Cribl to listen on port `10080` for HTTP (default). Set `authToken` to `myToken42`.
2. Send a payload to your Cribl Edge receiver.

Cribl Endpoint – Single Event

Cribl Single Event Example:

```
curl -k http://<myCriblHost>:10080/cribl/_bulk -H 'Authorization: myToken42' -d '{'
```

Cribl Endpoint – Multiple Events

Cribl Endpoint - Multiple Events

```
curl -k http://<myCriblHost>:10080/cribl/_bulk -H 'Authorization: myToken42' -d $'·
```

Splunk HEC Event Endpoint

Splunk HEC Event Endpoint

```
curl -k http://<myCriblHost>:10080/services/collector/event -H 'Authorization: myTo
```

```
curl -k http://<myCriblHost>:10080/services/collector -H 'Authorization: myToken42
```

For Splunk HEC, the token specification can be either `Splunk <token>` or `<token>`.

Cribl.Cloud – Single Event

1. Generate and copy a token in your Cribl.Cloud instance's HTTP Source > **General Settings**.
2. From the command line, use `https`, your Cribl.Cloud portal's **Ingest Endpoint** and port, and the token's value:

Cribl.Cloud – Single Event

```
curl -k https://default.main-<Your-Org-ID>.cribl.cloud:10080/cribl/_bulk -H 'Autho
```

With a Cribl.Cloud Enterprise plan, generalize the above URL's `default.main` substring to `<group-name>.main` when sending to other Fleets.

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

If one or more auth tokens are configured and enabled, Cribl Edge logs requests and events for each enabled auth token individually. Since the tokens themselves are redacted for security, Cribl Edge logs the initial text of the token description to help you identify which token a given log is for.

If no auth token is configured and enabled, Cribl Edge simply logs overall statistics about incoming requests and ingested events.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

10.11. Raw HTTP/S

Cribl Edge supports receiving raw HTTP data. The Raw HTTP Source listens on a specific port, captures every HTTP request to that port, and creates a corresponding event that it pushes to its configured Event Breakers.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **YES**

This Source supports gzip-compressed inbound data when the `Content-Encoding: gzip` connection header is set.

Configuring Cribl Edge to Receive Raw HTTP Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Pull** >] **Raw HTTP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Pull** >] **Raw HTTP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Raw HTTP Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the address to bind on. Defaults to `0.0.0.0` (all addresses).

Port: Enter the port number to listen on.

Authentication Settings

Auth tokens: Shared secrets to be provided by any client. Click **Generate** to create a new secret. If empty, permits open access.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

> In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.
>
> You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. This setting affects how the Source handles the `__srcIpPort` field.

Capture request headers: Toggle this to `Yes` to add request headers to events, in the `__headers` field.

Allowed URI paths: List of URI paths accepted by this input. Supports wildcards, e.g., `/api/v*/hook`. Defaults to `*`, which allows all paths.

Allowed HTTP methods: List of HTTP methods accepted by this input. Supports wildcards, e.g., `P*, GET`. Defaults to `*`, which allows all methods.

Max active requests: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.

Activity log sample rate: Determines how often request activity is logged at the `info` level. The default `100` value logs every 100th value; a `1` value would log every request; a `10` value would log every 10th request; etc.

Max requests per socket: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing Connection Reuse Against Request Distribution below.

Socket timeout (seconds): How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.

Request timeout (seconds): How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.

Keep-alive timeout (seconds): After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

The longer the **Keep-alive timeout**, the more Cribl Edge will reuse connections. The shorter the timeout, the closer Cribl Edge gets to creating a new connection for every request. When request frequency is high, you can use longer timeouts to reduce the number of connections created, which mitigates the associated cost.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields for this Source:

- `__channel`
- `__headers` – Automatically includes any headers sent with request.
- `__inputId`
- `__srcIpPort` – See details below.

Overriding `__srcIpPort` with Client IP/Port

The `__srcIpPort` field's value contains the IP address and (optionally) port of the HTTP client sending data to this Source.

When any proxies (including load balancers) lie between the HTTP client and the Source, the last proxy adds an `X-Forwarded-For` header whose value is the IP/port of the original client. With multiple proxies, this header's value will be an array, whose first item is the original client IP/port.

If `X-Forwarded-For` is present, and **Advanced Settings** > **Enable proxy protocol** is set to `No`, the original client IP/port in this header will override the value of `__srcIpPort`.

If **Enable proxy protocol** is set to `Yes`, the `X-Forwarded-For` header's contents will **not** override the `__srcIpPort` value. (Here, the upstream proxy can convey the client IP/port without using this header.)

10.12. Metrics

Cribl Edge supports receiving metrics in these wire formats/protocols: StatsD, StatsD Extended, and Graphite. Automatic protocol detection happens on the first line received over a TCP connection or a UDP packet. Lines not matching the detected protocol are dropped.

Type: **Push** | TLS Support: **No** | Event Breaker Support: **No**

Cribl Edge Workers support System Metrics only when running on Linux, not on Windows.

Configuring Cribl Edge to Receive Metrics

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Metrics**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [[**Push** >] **Metrics**. Next, click **New Source** to open a **New Source** modal that provides the options below.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP to listen to. Defaults to `0.0.0.0`.

UDP port: Enter the UDP port number to listen on. Not required if listening on TCP.

TCP port: Enter the TCP port number to listen on. Not required if listening on UDP.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP senders. Cribl also recommends tuning the OS UDP buffer size.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (TCP Only)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

IP allowlist regex: Regex matching IP addresses that are allowed to send data. Defaults to `.*` (i.e., all IPs.)

**Max buffer size (events) **: Maximum number of events to buffer when downstream is blocking. Defaults to `1000`.

UDP socket buffer size (bytes): Optionally, set the SO_RCVBUF socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Min: `256`. Max: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__srcIpPort`
- `__metricsInType`

Metric Event Schema and Destination Support

Metric data is read into the following event schema:

```
_metric - the metric name
_metric_type - the type of the metric (gauge, counter, timer)
_value - the value of the metric
_time - metric_time or Date.now()/1000
dim1 - value of dimension1
dim2 - value of dimension2
....
```

Cribl Edge places sufficient information into a field called `__criblMetric` to enable these events to be properly serialized out to any metric outputs (independent of the input type).

The following Destinations natively support the `__criblMetric` field:

- Splunk
- Splunk HEC
- InfluxDB
- Statsd

- Statsd Extended
- Graphite

Data Format/Protocol Examples

StatsD

Format: `MetricName:value|type`

StatsD Example

```
metric1:100|g
metric2:200|ms
metric.dot.3:300.16|c
```

See the StatsD repo.

StatsD Extended

Format: `MetricName:value|type|#dim=value,dim2=value`

StatsD Extended Example

```
metric1:100|g|#dim1:val1,dim2:val2,dim3:val3
metric2:200|ms|#dim1:val1,dim2:val2,dim3:val3
metric.dot.3:300.16|c|#dim1:val1,dim2:val2,dim3:val3
```

Graphite

Format: `MetricName[;dim1=val1[;dim2=val2]] value time`

Graphite Example with Dimensions

```
metric1;dim1=val1;dim2=val2 100 9999
metric2;dim1=val1;dim2=val2 200 9999
metric.dot.3;dim1=val1;dim2=val2 300.16 9999.16
```

Graphite Example without Dimensions

```
metric1 100 9999
metric2 200 9999
metric.dot.3 300.16 9999.16
```

See the Graphite (also known as Carbon) plaintext protocol.

10.13. NetFlow

Cribl Edge supports receiving NetFlow v5 data via UDP.

Type: **Push** | TLS Support: **No** | Event Breaker Support: **No**

This Source ingests NetFlow records similarly to how it ingests events from other upstream senders: fields are broken out, and the message header is included with each record. If you prefer to render NetFlow data as metrics, use a pre-processing Pipeline or a Route.

Configuring Cribl Edge to Receive NetFlow Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **+ Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **NetFlow**. Next, click either **+ Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **NetFlow**. Next, click **Add Source** to open a **New Source** modal that provides the options below.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP sender. Cribl also recommends tuning the OS UDP buffer size.

General Settings

Input ID: Enter a unique name to identify this NetFlow Source definition.

Address: Enter the hostname/IP to listen for NetFlow data. For example: `localhost`, `0.0.0.0`, or `::`.

Port: Enter the port number.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.
>
> For more about PQ modes, see Always On versus Smart Mode.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Cribl Edge has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (that is, all IPs).

IP denylist regex: Regex matching IP addresses whose messages you want this Source to ignore. Defaults to `^$` (that is, every specific IP address in the list). This takes precedence over the allowlist.

UDP socket buffer size (bytes): Optionally, set the `SO_RCVBUF` socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Minimum: `256`. Maximum: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

What Fields to Expect

The NetFlow Source does minimal processing of the incoming UDP messages to remain consistent with the internal Cribl event model. For each UDP message received on the socket, you can expect an event with the following fields. We have organized these fields into categories to make them easier to grasp; the categories are our own, and not from the NetFlow specifications. The field definitions are mostly copied from Cisco NetFlow documentation.

General

- `_time`: The UNIX timestamp (in seconds) at which the message was received by Cribl Edge.
- `source`: A string in the form `udp|<remote IP address>|<remote port>`, indicating the remote sender.
- `host`: The hostname of the machine running Cribl Edge that ingested this event.
- `inputInt`: SNMP index of input interface; always set to zero.
- `outputInt`: SNMP index of output interface.
- `protocol`: IP protocol type (for example, TCP = 6; UDP = 17) of the observed network flow.
- `tcpFlags`: Cumulative logical OR of TCP flags in the observed network flow.
- `tos`: IP type of service; switch sets it to the ToS of the first packet of the flow.
- `icmpType`: Type of the ICMP message. Only present if the protocol is ICMP.
- `icmpCode`: Code of the ICMP message. Only present if the protocol is ICMP.

Because this Source reads input data directly from bytes in a compact format, there is nothing suitable to put in a `_raw` field, and the Source does not add a `_raw` field to events.

Flow Source and Destination

- `srcAddr`: Source IP address; in case of destination-only flows, set to zero.
- `dstAddr`: Destination IP address.

- `nextHop`: IP address of next hop router.
- `srcPort`: TCP/UDP source port number.
- `dstPort`: TCP/UDP destination port number. If this is an ICMP flow, this field is a combination of ICMP type and ICMP code, which are broken out separately as `icmpType` and `icmpCode` fields.
- `srcMask`: Source address prefix mask bits.
- `dstMask`: Destination address prefix mask bits.
- `srcAs`: Autonomous system number of the source, either origin or peer.
- `dstAs`: Autonomous system number of the destination, either origin or peer.

Flow Statistics

- `packets`: Packets in the flow.
- `octets`: Total number of Layer 3 bytes in the packets of the flow.
- `startTime`: System uptime, in milliseconds, at the start of the flow.
- `endTime`: System uptime, in milliseconds, at the time the last packet of the flow was received.
- `durationMs`: `endTime` minus `startTime`.

Header

The following are subfields within the `header` field:

- `count`: Number of flows exported in this flow frame (protocol data unit, or PDU).
- `sysUptimeMs`: Current time in milliseconds since the export device booted
- `unixSecs`: Current seconds since 0000 UTC 1970.
- `unixNsecs`: Residual nanoseconds since 0000 UTC 1970.
- `flowSequence`: Sequence counter of total flows seen.
- `engineType`: Type of flow switching engine.
- `engineId`: ID number of the flow switching engine.
- `samplingMode`: The first two bits of what Cisco NetFlow documentation calls `SAMPLING_INTERVAL`. These bits specify a sampling mode.
- `samplingInterval`: The remaining 14 bits of what Cisco NetFlow documentation calls `SAMPLING_INTERVAL`. These bits specify a sampling interval.
- `version`: NetFlow export format version number.

Also, the internal fields listed below will be present.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields accessible for this Source:

- `__bytes`
- `__inputId`
- `_time`

UDP Tuning

Incoming UDP traffic is put into a buffer by the Linux kernel. Cribl will drain from that buffer as resources are available. At lower throughput levels, and with plenty of available processes, this isn't an issue. As you scale up, however, the default size of that buffer may be too small.

You can check the current buffer size with:

```
$ sysctl net.core.rmem_max
```

A typical value of about 200 KB is far too small for an even moderately busy syslog server. You can check the health of UDP with the following command. Check the `packet receive errors` line.

```
$ netstat -su
```

If `packet receive errors` is more than zero, you have lost events, which is a particularly serious problem if the number of errors is increasing rapidly. This means you need to increase your `net.core.rmem_max` setting (see earlier).

You can update the live settings, but you'll also need to change the boot-time setting so next time you reboot everything is ready to roll.

Live change, setting to 25 MB:

```
$ sysctl -w net.core.rmem_max=26214400
net.core.rmem_max = 26214400
$ sysctl -w net.core.rmem_default=26214400
net.core.rmem_default = 26214400
```

For the permanent settings change, add the following lines to `/etc/sysctl.conf`:

```
net.core.rmem_max=26214400
net.core.rmem_default=26214400
```

We recommend you track a few things related to UDP receiving:

- The `netstat -su` command, watching for errors.
- The **Status** tab in the UDP (Raw) Source. In particular, watch for dropped messages. They could indicate you need a bigger buffer under **Advanced Settings** (default: 1000 events). They could also indicate your Worker is encountering pressure further down the Pipeline.
- Especially if you increase your kernel receive buffer as above, watch your Worker processes' memory usage.

10.14. OpenTelemetry (OTel)

Cribl Edge supports receiving trace and metric events from OTLP-compliant senders. (Cribl plans to add support for log events as more components of the OpenTelemetry protocol's logs support graduate to stable status.)

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

Supported and Unsupported Input Data

In Cribl Edge 4.0.3 and later, the Open Telemetry Source supports receiving compressed inbound data (with DEFLATE or gzip compression), as well as uncompressed data.

In Cribl Edge 4.1 and later, this Source supports receiving telemetry data over either of the transports that the OpenTelemetry Protocol (OTLP) describes: gRPC or HTTP. OTLP defines Protocol buffer (Protobuf) schemas for its payloads (requests and responses). With the HTTP transport, this Source supports Binary Protobuf payload encoding, but currently does not support JSON Protobuf.

The OpenTelemetry Project's Data Sources documentation provides these hierarchical definitions of Cribl Edge's supported trace and metric event types:

- A **trace** tracks the progression of a single request.
- Each trace is a tree of **spans**.
- A span object represents the work being done by the individual services, or components, involved in a request as that request flows through a system.
- A **metric** provides aggreggated statistical information.
- A metric contains individual measurements called **data points**.

Configuring an OTel Source

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **OpenTelemetry**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **OpenTelemetry**. Next, click **New Source** to open a

New Source modal that provides the options below.

The sections described below are spread across several tabs. Click the tab links at left, or the **Next** and **Prev** buttons, to navigate among tabs. Click **Save** when you've configured your Source.

General Settings

Input ID: Unique ID for this Source. E.g., `OTel042`.

Address: Enter the hostname/IP to listen on. Defaults to `0.0.0.0` (all addresses, IPv4 format).

Port: By default, OTel applications send output to port `4317` when using the gRPC protocol, and port `4318` when using HTTP. This setting defaults to `4317` – you must change it if you set **Protocol** (below) to `HTTP`, or you want Cribl Edge to collect data from an OTel application that is using a different port.

Optional Settings

Protocol: Use the drop-down to choose the protocol matching the data you will ingest: `gRPC` (the default), or `HTTP`.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Enter the bearer token that must be included in the authorization header.
- **Auth token (text secret)**: Provide an HTTP token referenced by a secret. Select a stored text secret in the resulting drop-down, or click **Create** to configure a new secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: Provide username and password credentials referenced by a secret. Select a stored text secret in the resulting **Credentials secret** drop-down, or click **Create** to configure a new secret.

TLS Settings (Server Side)

In OTel terminology, your Cribl Edge OTel Source will receive OTel data from a **Collector** running on a local **agent**. In Cribl Edge's terminology, the Collector is the **client** and the OTel Source is the **server**. This is why this Source's UI identifies the Source's TLS Settings as "server-side."

For more about this client-server relationship, see the TLS Configuration Example below.

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs (mTLS). Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (for example, the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below,

> use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Advanced Settings always displays **Extract spans**, **Extract metrics**, and **Environment**:

The **Extract spans** and **Extract metrics** settings are unique to OTel. Their default `No` settings allow Cribl Edge to essentially function as a bump on the wire, generating a single event for each incoming OTel event. This can be useful when, for example, you want to send whole OTel events to persistent storage.

- **Extract spans**: Toggle to `Yes` if you want Cribl Edge to generate an individual event for each span. (Recall that traces contain multiple spans.)
- **Extract metrics**: Toggle to `Yes` if you want Cribl Edge to generate an individual event for each data point. (Recall that OTel metric events contain multiple data points.)
- **Environment**: Optionally, specify a single Git branch on which to enable this configuration. If this field is empty, the config will be enabled everywhere.

When **General Settings** > **Protocol** is set to `gRPC`, the UI displays one additional setting:

- **Max active connections**: Maximum number of active connections allowed per Worker Process. Defaults to `1000`. Set a lower value if connection storms are causing the Source to hang. Set `0` for unlimited connections.

When **General Settings** > **Protocol** is set to `HTTP`, the UI displays five additional settings:

- **Max active requests**: Maximum number of active requests allowed for this Source, per Worker Process. Defaults to `256`. Enter `0` for unlimited.
- **Max requests per socket**: The maximum number of requests Cribl Edge should allow on one socket before instructing the client to close the connection. Defaults to `0` (unlimited). See Balancing

Connection Reuse Against Request Distribution below.

- **Socket timeout (seconds)**: How long Cribl Edge should wait before assuming that an inactive socket has timed out. The default `0` value means wait forever.
- **Request timeout (seconds)**: How long to wait for an incoming request to complete before aborting it. The default `0` value means wait indefinitely.
- **Keep-alive timeout (seconds)**: After the last response is sent, Cribl Edge will wait this long for additional data before closing the socket connection. Defaults to `5` seconds; minimum is `1` second; maximum is `600` seconds (10 minutes).

Balancing Connection Reuse Against Request Distribution

Max requests per socket allows you to limit the number of HTTP requests an upstream client can send on one network connection. Once the limit is reached, Cribl Edge uses HTTP headers to inform the client that it must establish a new connection to send any more requests. (Specifically, Cribl Edge sets the HTTP `Connection` header to `close`.) After that, if the client disregards what Cribl Edge has asked it to do and tries to send another HTTP request over the existing connection, Cribl Edge will respond with an HTTP status code of `503 Service Unavailable`.

Use this setting to strike a balance between connection reuse by the client, and distribution of requests among one or more Edge Node processes by Cribl Edge:

- When a client sends a sequence of requests on the same connection, that is called connection reuse. Because connection reuse benefits client performance by avoiding the overhead of creating new connections, clients have an incentive to **maximize** connection reuse.
- Meanwhile, a single process on that Edge Node will handle all the requests of a single network connection, for the lifetime of the connection. When receiving a large overall set of data, Cribl Edge performs better when the workload is distributed across multiple Edge Node processes. In that situation, it makes sense to **limit** connection reuse.

There is no one-size-fits-all solution, because of variation in the size of the payload a client sends with a request and in the number of requests a client wants to send in one sequence. Start by estimating how long connections will stay open. To do this, multiply the typical time that requests take to process (based on payload size) times the number of requests the client typically wants to send.

If the result is 60 seconds or longer, set **Max requests per socket** to force the client to create a new connection sooner. This way, more data can be spread over more Edge Node processes within a given unit of time.

For example: Suppose a client tries to send thousands of requests over a very few connections that stay open for hours on end. By setting a relatively low **Max requests per socket**, you can ensure that the same work is done over more, shorter-lived connections distributed between more Edge Node processes, yielding better performance from Cribl Edge.

A final point to consider is that one Cribl Edge Source can receive requests from more than one client, making it more complicated to determine an optimal value for **Max requests per socket**.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

TLS Configuration Example

Here's a simple example for using TLS to secure the communication between an OpenTelemetry client and your Cribl Edge OTel Source.

1. Choose or generate a certificate and key. If you need to generate a certificate/key pair, you can adapt the following OpenSSL command:

   ```
   openssl req -nodes -new -x509 -newkey rsa:2048 -keyout myKey.pem -out myCert.p
   ```

   This example command will generate both a self-signed cert named `myCert.pem` (certified for 420 days), and an unencrypted, 2048-bit RSA private key named `myKey.pem`.
2. Configure the **TLS Settings (Server Side)**. Toggle **Enabled** to `Yes`, then:
   - Enter the appropriate values in the **Certificate name**, **Private key path**, and **Certificate path** fields. A **Create** link is available if you need a new certificate, and **Certificate name** also works as a drop-down to allow you to choose from any existing certificates.
   - Leave the **CA certificate path** field empty.
   - Leave **Authenticate client (mutual auth)** toggled to `No`.
3. Configure the OTel client. See the OTel Collector TLS Configuration Settings README for an explanation of the relevant settings. The config file might be named `otel-config.yaml`, `otel-local-config.yaml`, or just `config.yaml`, depending on your environment. This YAML file will have an `exporters` section, which you must edit to include an `otlp` sub-section, as follows:
   - Add an `endpoint` whose value is the IP address of either (a) the Cribl Edge Edge Node on which your OTel Source is running, or (b) the IP address of the load balancer for the relevant Fleet. In the example snippet below, this is the `<Cribl_IP_address>`. Specify the port on which Cribl Edge's OTel Source is listening; port `4317` is the default.
   - Set `tls` > `insecure` to `false`. This matches your setting **TLS Settings (Server Side)** > **Enabled** to `Yes` on the Cribl Edge OTel Source.

- Set `tls` > `insecure_skip_verify` to `true`. This matches your setting **TLS Settings (Server Side)** > **Authenticate client (mutual auth)** to `No` on the Cribl Edge OTel Source. Setting `insecure_skip_verify` to `true` is also required if you're using a self-signed certificate.

Here's how the section you edited should look:

```
exporters:
  otlp:
    endpoint: "https://<Cribl_IP_address>:4317"
    tls:
      insecure: false
      insecure_skip_verify: true
```

10.15. SNMP Trap

Cribl Edge supports receiving data from SNMP Traps.

Type: **Push** | TLS Support: **NO** | Event Breaker Support: **No**

Configuring Cribl Edge to Receive SNMP Traps

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **SNMP Trap**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **SNMP Trap**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge, except in Cribl.Cloud, ships with an SNMP Trap Source preconfigured to listen on port `9162`. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Address to bind on. Defaults to `0.0.0.0` (all addresses).

UDP Port: Port on which to receive SNMP traps. Defaults to `162`.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP senders. Cribl also recommends tuning the OS UDP buffer size.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication Settings

SNMPv3 authentication provides secure access to devices through optional user authentication and decryption of incoming data packets. Cribl Edge lets you choose any of the following levels of security:

- User name checking without authentication.
- User authentication without privacy.
- User authentication with privacy.

SNMPv3 authentication: Toggle to `Yes` to enable SNMPv3 authentication and reveal authentication parameters. Defaults to `No`.

Allow unmatched traps: Toggle to `Yes` to pass through traps that don't match any of the configured users. Cribl Edge will not attempt to authenticate or decrypt these traps. When toggled to `No` (default), Cribl Edge drops traps without a correctly configured user name.

v3 Name: Enter the SNMPv3 user name (required). Multiple users are supported.

You must configure at least one user to enable SNMPv3 authentication. If desired, you can enter separate user credentials for each SNMPv3 user. The authentication protocol, authentication key, privacy protocol, and privacy key can be unique for each user.

For authentication to succeed, the user name and authentication protocol and key must match in the Source and the sending device configuration.

For decryption to succeed, the user name, authentication protocol and key, and the privacy protocol and key must match in the Source and the sending device configuration.

Authentication protocol: Select the authentication protocol required for your use case. If you select `None`, Cribl Edge will only check the user name without authenticating the user or decrypting the data. Otherwise, Cribl Edge uses the authentication protocol you specify and the key you provide to authenticate the user.

Options include:

- `None`
- `MD5`
- `SHA`

- `SHA224`
- `SHA256`
- `SHA384`
- `SHA512`

v3 authentication key: Enter the authentication key.

Choosing an authentication protocol also allows you to select an optional privacy protocol to decrypt incoming packets.

Cribl Edge logs authentication failures at the `debug` level.

Privacy protocol: Select the privacy protocol required for your use case. If you select `None`, Cribl Edge authenticates the user without decrypting the incoming data. Otherwise, Cribl Edge uses the privacy protocol you specify and the key you provide to decrypt the data.

Options include:

- `None`
- `DES`
- `AES`
- `AES256b (Blumenthal)`
- `AES2556r (Reeder)`

For every successfully decrypted trap, Cribl Edge adds `__didDecrypt: true` to the Event. Cribl Edge logs decryption failures at the `debug` level.

v3 privacy key: Enter the privacy key.

Add user: Click to add a new set of user credentials for SNMPv3 authentication.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always On`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

IP allowlist regex: Regex matching IP addresses that are allowed to send data. Defaults to `.*`, i.e., all IPs.

Max buffer size (events): Maximum number of events to buffer when downstream is blocking. Defaults to `1000`.

UDP socket buffer size (bytes): Optionally, set the SO_RCVBUF socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Min: `256`. Max: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Fields for this Source:

- `__didDecrypt`: Set to `true` for every trap that is successfully decrypted.
- `__final`
- `__inputId`
- `__snmpRaw`: Buffer containing Raw SNMP packet
- `__snmpVersion`: Acceptable values are `0`, `2` , or `3`. These respectively indicate SNMP v1, v2c, and v3.
- `__srcIpPort` : In this particular Source, this field uses a pipe (|) symbol to separate the source IP address and the port, in this format: `event.__srcIpPort = ${rInfo.address}|${rInfo.port};`
- `_time`

Considerations for Working with SNMP Trap Data

- It's possible to work with SNMP metadata (i.e., we'll decode the packet). Options include dropping, routing, etc.
- SNMP packets can be forwarded to other SNMP destinations. However, the contents of the incoming packet **cannot** be modified – i.e., we'll forward the packets verbatim as they came in.
- SNMP packets can be forwarded to non-SNMP destinations (e.g., Splunk, Syslog, S3, etc.).
- Non-SNMP input data **cannot** be sent to SNMP destinations.

10.16. Syslog

Cribl Edge supports receiving syslog data, whether structured according to RFC 3164 or RFC 5424. This Source supports message-length prefixes according to RFC 5425 or RFC 6587.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

For details on how to replace your syslog server with Cribl Edge, see Syslog Best Practices.

Configuring Cribl Edge to Receive Data over Syslog

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Syslog**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Syslog**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with a Syslog Source preconfigured to listen for both UDP and TCP traffic on Port 9514. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this Syslog Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter the hostname/IP on which to listen for data., E.g. `localhost` or `0.0.0.0`.

UDP port: Enter the UDP port number to listen on. Not required if listening on TCP.

The maximum supported inbound UDP message size is 16,384 bytes.

TCP port: Enter the TCP port number to listen on. Not required if listening on UDP.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP senders. Cribl also recommends tuning the OS UDP buffer size.

Optional Settings

Fields to keep: List of fields from source data to retain and pass through. Supports wildcards. Defaults to `*` wildcard, meaning keep all fields. Fields **not** specified here (by wildcard or specific name) will be removed from the event.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (TCP Only)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries,

the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data.

Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

Single msg per UDP: Enable this to treat received UDP packet data as a full syslog message. With the `No` default, Cribl Edge will treat newlines within the packet as event delimiters.

Octet count framing: Toggle to `Yes` if messages are prefixed with a byte length, according to RFC 5425 or RFC 6587. The default setting (`No`) applies non-transparent framing using `\n` as the delimiter. The framing

method utilized by this input will be applied to all events received by this input. Therefore, if your syslog devices use a mixture of framing types (non-transparent vs. octet count), you will need to use a separate Syslog Source for each framing type. Additional inputs will necessitate separate ports.

Allow non-standard app name: Toggle to `Yes` to allow hyphens to appear in an RFC 3164–formatted Syslog message's `TAG` section. For details, see TAG Section Processing.

Enable TCP load balancing: Toggle to `Yes` if you want the Source to load balance traffic across all Worker Processes, as explained below. (This setting is available only in Cribl Stream deployed in Distributed mode.)

IP whitelist regex: Regex matching IP addresses that are allowed to send data. Defaults to `.*` (i.e., all IPs).

Max buffer size (events): Maximum number of events to buffer when downstream is blocking. The buffer is only in memory. (This setting is applicable only to UDP syslog.) Events dropped because they exceed this buffer are logged as dropped in `stats` messages.

UDP socket buffer size (bytes): Optionally, set the SO_RCVBUF socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Min: `256`. Max: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Default timezone: Timezone to assign to timestamps that omit timezone info. Accept the default `Local` value, or use the drop-down list to select a specific timezone by city name or GMT/UTC offset.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

TCP Load Balancing

This feature is available only for Cribl Stream deployments that are in Distributed mode.

When the Syslog Source receives a high volume of data over TCP, a single Worker Process can place excessive CPU load on its Cribl Stream Worker Node, while remaining Worker Processes on the same node mostly sit idle. This undesirable condition is called "TCP pinning" because the high volume of traffic "pins" a single TCP connection.

To alleviate TCP pinning, try toggling **Advanced Settings** > **Enable TCP load balancing** on. The Syslog Source will then load balance traffic across all Worker Processes.

When TCP load balancing is enabled, the Worker Node forks a special Worker Process dedicated to load balancing; creates sockets for communication between the load balancer and the regular Worker Processes; and, makes TCP load balancing metrics available.

The Load Balancing Process

The Worker Node forks a new, special **load balancer** Worker Process that distributes incoming syslog data among the other Worker Processes. For customer-managed and hybrid Fleets, consider reducing the Process count by 1, if possible. This will leave a core free to run the load balancer process.

If you want to verify that the load balancer Worker Process exists, teleport into the relevant Worker and in the **Worker Processes** tab you will see a Worker Process with `LB` in its name.

To view logs from the Leader UI, navigate to **Monitoring** > **Logs** > `<Worker Node ID>` > **Load Balancer**.

Worker Process Socket Files

The load balancer Worker Process sends data to the regular Worker Processes over inter-process communication (IPC) sockets. This means that in the `state` directory, there is a socket file for each load balancer Worker Process and each Worker Process.

By default, the socket files are linked to `state`, from the Cribl `tmp` directory. For customer-managed and hybrid Worker Groups, you can specify a directory other than `tmp` for these links. Cribl recommends that you do **not** use this option, but if you find it necessary, navigate to **Group Settings** > **General Settings** > **Sockets** to configure as desired.

TCP Load Balancing Metrics

Metrics specific to TCP load balancing become available.

- `lb.bytes_out` – Total bytes processed by the load balancer Worker Process, by Source.
- `lb.writable_sockets` – Total unblocked Worker Process sockets, by Source.
- `lb.blocked_sockets` – Total blocked Worker Process sockets, by Source.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

TAG Section Processing

Cribl Edge will try to parse `appname` out of a syslog message's `TAG` section, even when the message is not RFC 5424–formatted.

This practice means that the `TAG` section can contain any alphanumeric character or an underscore (`_`). When the **Allow–non-standard app name** option is enabled, Cribl Edge will also process hyphens that appear in an RFC 3164–formatted Syslog message's `TAG` section. If Cribl Edge encounters a hyphenated `appname`, it will continue processing to find `procid`. (This setting has no effect on RFC 5424–formatted messages.)

If the `TAG` section contains any non-alphanumeric character, Cribl Edge will treat that character as the termination of the `TAG` section, and as the starting character of the `CONTENT` section.

CONTENT Section Processing

Cribl Edge will try to parse `procid` from the beginning of the `CONTENT` section if this section is directly after the `TAG` section, and if it either follows a `:` or is surrounded by `[]`. This process occurs regardless of the **Allow–non-standard app name** setting.

What Fields to Expect

To create fields, the Syslog Source requires messages to comply with RFC 3164 or RFC 5424. If messages sent to the Source comply with neither of these RFCs, you'll see the following fields:

- `_raw`: Contains the whole message, but no other fields broken out.
- `_time`: Contains the time the message was received by Cribl Edge.
- `__srcIpPort <udp|tcp>:<port>`: See Internal Fields.
- `__syslogFail`: See Internal Fields.
- `host`: IP address of the device sending the incoming message.

If messages sent to the Source comply with either RFC 3164 or RFC 5424, fields that the RFC deems guaranteed will always be there, but fields deemed optional might or might not be. Once Cribl Edge parses the required fields and any optional fields, what remains is the actual message.

To see this in real life, install the `cribl-syslog-input` Pack and preview the `RFC5424-RFC3164.log` sample file.

RFC Name for Field	Cribl Name for Field	Guaranteed?	Notes
PRI	facility, facilityName, severity, severityName	Yes	PRI is a bitwise representation of Facility and Severity, which is how Cribl Edge can derive values for all four of its fields.
TIMESTAMP	_time	Yes	This field's format differs depending on which RFC the messages adhere to. Cribl Edge converts it to UNIX epoch time.
HOSTNAME	host	Yes	
APP-NAME	appname	No	
PROCID	procid	No	
MSGID	msgid	No	
STRUCTURED-DATA	structuredData	No	
MSG	message	Yes	

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but are accessible and Functions can use them to make processing decisions.

Fields for this Source:

- `__inputId`
- `__srcIpPort <udp|tcp>:<port>`: Identifies the port of the sending socket.
- `__syslogFail: true` for data that fails RFC 3164/5424 validation as syslog format.

`stats` Log Message

In Cribl Edge 4.2.2 and later, `stats` log messages report the number of events received, buffered, or dropped for exceeding the maximum Cribl buffer size. By default, these messages are logged every 60 seconds. These values also appear in **Sources** > **Syslog** > <Source_name> > **Status**, in the **UDP** row.

UDP Tuning

Incoming UDP traffic is put into a buffer by the Linux kernel. Cribl will drain from that buffer as resources are available. At lower throughput levels, and with plenty of available processes, this isn't an issue. As you scale up, however, the default size of that buffer may be too small.

You can check the current buffer size with:

```
$ sysctl net.core.rmem_max
```

A typical value of about 200 KB is far too small for an even moderately busy syslog server. You can check the health of UDP with the following command. Check the `packet receive errors` line.

```
$ netstat -su
```

If `packet receive errors` is more than zero, you have lost events, which is a particularly serious problem if the number of errors is increasing rapidly. This means you need to increase your `net.core.rmem_max` setting (see earlier).

You can update the live settings, but you'll also need to change the boot-time setting so next time you reboot everything is ready to roll.

Live change, setting to 25 MB:

```
$ sysctl -w net.core.rmem_max=26214400
net.core.rmem_max = 26214400
$ sysctl -w net.core.rmem_default=26214400
net.core.rmem_default = 26214400
```

For the permanent settings change, add the following lines to `/etc/sysctl.conf`:

```
net.core.rmem_max=26214400
net.core.rmem_default=26214400
```

We recommend you track a few things related to UDP receiving:

- The `netstat -su` command, watching for errors.
- The **Status** tab in the Syslog Source. In particular, watch for dropped messages. They could indicate you need a bigger buffer under **Advanced Settings** (default: 1000 events). They could also indicate

your Worker is encountering pressure further down the Pipeline.

- Especially if you increase your kernel receive buffer as above, watch your Worker processes' memory usage.

Troubleshooting Resources

Cribl University offers an Advanced Troubleshooting > Source Integrations: Syslog short course. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

10.17. TCP JSON

Cribl Edge can receive newline-delimited JSON data over TCP.

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **No**

Configuring Cribl Edge to Receive TCP JSON Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **TCP JSON**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **TCP JSON**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with a TCP JSON Source preconfigured to listen on Port 10070. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this TCP JSON Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for TCP JSON data. E.g., `localhost` or `0.0.0.0`.

Port: Enter the port number to listen on.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Use this default option to enter the shared secret that clients must provide in the `authToken` header field. Exposes an **Auth token** field for this purpose. (If left blank, unauthenticated access will be permitted.) A **Generate** link is available if you need a new secret.

- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the `authToken` header field value described above. The secret can reside in Cribl Edge's internal secrets manager or (if enabled) in an external KMS. A **Create** link is available if you need a new secret.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event, using Eval-like functionality.

Name: Field name.

Value: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2.

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (i.e., all IPs).

Max active connections: Maximum number of active connections allowed per Worker Process. Defaults to `1000`. Set a lower value if connection storms are causing the Source to hang. Set `0` for unlimited connections.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and Functions can use them to make processing decisions.

Field for this Source:

- `__inputId`
- `__srcIpPort`

Format

Cribl Edge expects TCP JSON events in newline-delimited JSON format:

1. A header line. Can be empty – e.g., `{}`. If **authToken** is enabled (see above) it should be included here as a field called `authToken`. When `authToken` is **not** set, the header line is **optional**. In this case, the first line will be treated as an event if does not look like a header record.

In addition, if events need to contain common fields, they can be included here under `fields`. In the example below, `region` and `AZ` will be automatically added to all events.

2. A JSON event/record per line.

Sample TCP JSON Events

```
{"authToken":"myToken42", "fields": {"region": "us-east-1", "AZ":"az1"}}

{"_raw":"this is a sample event ", "host":"myHost", "source":"mySource", "fieldA":'
{"host":"myOtherHost", "source":"myOtherSource", "_raw": "{\"message\":\"Something
```

TCP JSON Field Mapping to Splunk

If a TCP JSON Source is routed to a Splunk destination, fields within the JSON payload are mapped to Splunk fields. Fields that do not have corresponding (native) Splunk fields become index-time fields. For example, let's assume we have a TCP JSON event as below:

```
{"_time":1541280341, "host":"myHost", "source":"mySource", "_raw":"this is a sample event ", "fieldA":"valueA"}
```

Here, `_time`, `host`, and `source` become their corresponding fields in Splunk. The value of `_raw` becomes the actual body of the event, and `fieldA` becomes an index-time field (`fieldA::`valueA``).

Examples

Testing TCP JSON In

This first example simply tests that data is flowing in through the Source:

1. Configure Cribl Edge to listen on port `10001` for TCP JSON. Set `authToken` to `myToken42`.
2. Create a file called `test.json` with the payload above.
3. Send it over to your Cribl Edge host: `cat test.json | nc <myCriblHost> 10001`

Cribl Edge to Cribl.Cloud

This second example demonstrates using TCP JSON to send data from one Cribl Edge instance to a downstream Cribl.Cloud instance. We assume that the downstream Cloud instance uses Cribl.Cloud's **default** TCP JSON Source configuration.

So all the configuration happens on the upstream instance's TCP JSON Destination. Replace the `<Your-Org-ID>` placeholder with the Org ID from your Cribl Cloud portal.

TCP JSON Destination Configuration

On the upstream Cribl Edge instance's Destination, set the following field values to match the target Cloud instance's defaults:

General Settings

Address: `default.main-<Your-Org-ID>.cribl.cloud` – you can simply copy/paste your Cribl.Cloud portal's **Ingest Endpoint** here. With a Cribl.Cloud Enterprise plan, generalize the `default.main` substring in this URL to `<group-name>.main` when sending to other Fleets.

Port: `10070`

TLS Settings (Client Side)

Enabled: `Yes`

Validate server certs: `Yes`

Periodic Logging

Cribl Edge logs metrics about incoming requests and ingested events once per minute.

These logs are stored in the `metrics.log` file. To view them in the UI, open the Source's **Logs** tab and choose **Worker Process X Metrics** from the drop-down, where **X** is the desired Worker process.

This kind of periodic logging helps you determine whether a Source is in fact still healthy even when no data is coming in.

10.18. TCP (Raw)

Cribl Edge supports receiving of data over TCP. (See examples and header options below.)

Type: **Push** | TLS Support: **YES** | Event Breaker Support: **YES**

Configuring Cribl Edge to Receive TCP Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **TCP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **TCP**. Next, click **New Source** to open a **New Source** modal that provides the options below.

Cribl Edge ships with a TCP Source preconfigured to listen on Port 10060. You can clone or directly modify this Source to further configure it, and then enable it.

General Settings

Input ID: Enter a unique name to identify this TCP Source definition. If you clone this Source, Cribl Edge will add `-CLONE` to the original **Input ID**.

Address: Enter hostname/IP to listen for raw TCP data. E.g., `localhost` or `0.0.0.0`.

Port: Enter port number.

Enable Header: Toggle to `Yes` to indicate that client will pass a header record with every new connection. The header can contain an `authToken`, and an object with a list of fields and values to add to every event. These fields can be used to simplify Event Breaker selection, routing, etc. Header format: `{ "authToken" : "myToken", "fields": { "field1": "value1", "field2": "value2" }}`.

- **Shared secret (authToken)**: Shared secret to be provided by any client (in `authToken` header field). Click **Generate** to create a new secret. If empty, unauthenticated access will be permitted.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Server Side)

Enabled defaults to `No`. When toggled to `Yes`:

Certificate name: Name of the predefined certificate.

Private key path: Server path containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`.

Passphrase: Passphrase to use to decrypt private key.

Certificate path: Server path containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

CA certificate path: Server path containing CA certificates (in PEM format) to use. Path can reference `$ENV_VARS`.

Authenticate client (mutual auth): Require clients to present their certificates. Used to perform mutual authentication using SSL certs. Defaults to `No`. When toggled to `Yes`:

- **Validate client certs**: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.
- **Common Name**: Regex that a peer certificate's subject attribute must match in order to connect. Defaults to `.*`. Matches on the substring after `CN=`. As needed, escape regex tokens to match literal characters. (For example, to match the subject `CN=worker.cribl.local`, you would enter: `worker\.cribl\.local`.) If the subject attribute contains Subject Alternative Name (SAN) entries, the Source will check the regex against all of those but ignore the Common Name (CN) entry (if any). If the certificate has no SAN extension, the Source will check the regex against the single name in the CN.

Minimum TLS version: Optionally, select the minimum TLS version to accept from connections.

Maximum TLS version: Optionally, select the maximum TLS version to accept from connections.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.

The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings. **Queue file path**: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

> You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Custom Command

In this section, you can pass the data from this input to an external command for processing before the data continues downstream.

Enabled: Defaults to `No`. When toggled to `Yes`:

Command: Enter the command that will consume the data (via `stdin`) and will process its output (via `stdout`).

Arguments: Click **Add Argument** to add each argument for the command. You can drag arguments vertically to resequence them.

Event Breakers

Event Breaker rulesets: A list of event breaking rulesets that will be applied to the input data stream before the data is sent through the Routes. Defaults to `System Default Rule`.

Event Breaker buffer timeout: How long (in milliseconds) the Event Breaker will wait for new data to be sent to a specific channel, before flushing out the data stream, as-is, to the Routes. Minimum `10` ms, default `10000` (10 sec), maxiumum `43200000` (12 hours).

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Enable Proxy Protocol: Toggle to `Yes` if the connection is proxied by a device that supports Proxy Protocol v1 or v2. When this setting is enabled, the `__srcIpPort` internal field will show the original source IP address and port. When it is disabled, the `__srcIpPort` field will show the IP address and port of the proxy that forwarded the connection.

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (i.e,. all IPs).

Max active connections: Maximum number of active connections allowed per Worker Process. Defaults to `1000`. Set a lower value if connection storms are causing the Source to hang. Set `0` for unlimited connections.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields accessible for this Source:

- `__inputId`
- `__srcIpPort`
- `__channel`

TCP Source Examples

Every new TCP connection may contain an **optional** header line, with an `authToken` and a list of fields and values to add to every event. To use the Cribl Edge Cloud sample, copy the `<token_value>` out of your Cribl Edge Cloud TCP Source.

Sample test.raw (on-prem) | Sample test.raw (Cribl.Cloud)

```
{"authToken":"myToken42", "fields": {"region": "us-east-1", "AZ":"az1"}}

this is event number 1
this is event number 2
```

Enabling the Example – Cribl Edge

1. Configure Cribl Edge to listen on port `7777` for raw TCP. Set `authToken` to `myToken42`.
2. Create a file called `test.raw`, with the payload above.
3. Send it over to your Cribl Edge host, using this command: `cat test.raw | nc <myCriblHost> 7777`

Enabling the Example – Cribl.Cloud

Use netcat with `--ssl` and `--ssl-verify`:

Command-line test

```
cat test.raw | nc --ssl --ssl-verify default.main-<Your-Org-ID>.cribl.cloud 10060
```

With a Cribl.Cloud Enterprise plan, generalize the above URL's `default.main` substring to `<group-name>.main` when sending to other Fleets.

10.19. UDP (Raw)

Cribl Edge supports receiving raw, unparsed data via UDP.

Type: **Push** | TLS Support: **NO** | Event Breaker Support: **NO**

Configuring Cribl Edge to Receive Raw UDP Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **+ Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **Raw UDP**. Next, click either **+ Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **Raw UDP**. Next, click **Add Source** to open a **New Source** modal that provides the options below.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP sender. Cribl also recommends tuning the OS UDP buffer size.

General Settings

Input ID: Enter a unique name to identify this raw UDP Source definition.

Address: Enter the hostname/IP to listen for raw UDP data. For example: `localhost`, `0.0.0.0`, or `::`.

Port: Enter the port number.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

In Cribl Edge 4.1 and later, Source-side PQ's default **Mode** is `Always on`, to best ensure events' delivery. For details on optimizing this selection, see Always On versus Smart Mode.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

Single msg per UDP: Toggle to `Yes` if each UDP message should be treated as an independent event. Leave set to the default `No` if the message should be broken on newlines to create multiple events.

Ingest raw bytes: Toggle to `Yes` to add a `__rawBytes` field to each event containing an array of the bytes received as the UDP message.

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (i.e, all IPs).

Max buffer size (events): Maximum number of events to buffer when downstream is blocking. The buffer is only in memory.

UDP socket buffer size (bytes): Optionally, set the SO_RCVBUF socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Min: `256`. Max: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

What Fields to Expect

The Raw UDP Source does minimal processing of the incoming UDP messages to remain consistent with the internal Cribl event model. For each UDP message received on the socket, you can expect an event with the following fields:

- `_raw`: Contains the UTF-8 representation of the entire message received (if **Single msg per UDP** is set to `Yes`), or of the given line that was split out of the message.
- `_time`: The UNIX timestamp (in seconds) at which the message was received by Cribl Edge.
- `source`: A string in the form `udp|<remote IP address>|<remote port>`, indicating the remote sender.
- `host`: The hostname of the machine running Cribl Edge that ingested this event.

Also, the internal fields listed below will be present.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields accessible for this Source:

- `__inputId`
- `__srcIpPort`
- `__rawBytes`: When **Ingest raw bytes** is set to `Yes`, this field will be an array containing the bytes of the UDP message.

UDP Tuning

Incoming UDP traffic is put into a buffer by the Linux kernel. Cribl will drain from that buffer as resources are available. At lower throughput levels, and with plenty of available processes, this isn't an issue. As you scale up, however, the default size of that buffer may be too small.

You can check the current buffer size with:

```
$ sysctl net.core.rmem_max
```

A typical value of about 200 KB is far too small for an even moderately busy syslog server. You can check the health of UDP with the following command. Check the `packet receive errors` line.

```
$ netstat -su
```

If `packet receive errors` is more than zero, you have lost events, which is a particularly serious problem if the number of errors is increasing rapidly. This means you need to increase your `net.core.rmem_max` setting (see earlier).

You can update the live settings, but you'll also need to change the boot-time setting so next time you reboot everything is ready to roll.

Live change, setting to 25 MB:

```
$ sysctl -w net.core.rmem_max=26214400
net.core.rmem_max = 26214400
$ sysctl -w net.core.rmem_default=26214400
net.core.rmem_default = 26214400
```

For the permanent settings change, add the following lines to `/etc/sysctl.conf`:

```
net.core.rmem_max=26214400
net.core.rmem_default=26214400
```

We recommend you track a few things related to UDP receiving:

- The `netstat -su` command, watching for errors.
- The **Status** tab in the UDP (Raw) Source. In particular, watch for dropped messages. They could indicate you need a bigger buffer under **Advanced Settings** (default: 1000 events). They could also

indicate your Worker is encountering pressure further down the Pipeline.

- Especially if you increase your kernel receive buffer as above, watch your Worker processes' memory usage.

11. Destinations

Cribl Edge can send data to various Destinations, including Splunk, Kafka, Kinesis, InfluxDB, Snowflake, Databricks, TCP JSON, and many others. Destinations can write data to either IPv4 or IPv6 addresses.

Destinations in the Edge ecosystem

Streaming Destinations

Destinations that accept events in real time are referred to as streaming Destinations:

- Amazon CloudWatch Logs
- Amazon Kinesis Streams
- Amazon MSK
- Amazon SQS
- Azure Data Explorer when in streaming mode
- Azure Event Hubs
- Azure Monitor Logs
- Azure Sentinel
- Confluent Cloud
- Cribl HTTP
- Cribl TCP
- CrowdStrike Falcon LogScale
- Datadog
- Elasticsearch
- Elastic Cloud

- Google Chronicle
- Google Cloud Logging
- Google Cloud Pub/Sub
- Grafana Cloud
- Graphite
- Honeycomb
- InfluxDB
- Kafka
- Loki
- New Relic Events
- New Relic Logs & Metrics
- OpenTelemetry (OTel)
- Prometheus
- SentinelOne DataSet
- SignalFx
- SNMP Trap
- Splunk HEC
- Splunk Load Balanced
- Splunk Single Instance
- StatsD
- StatsD Extended
- Sumo Logic
- Syslog
- TCP JSON
- Wavefront
- Webhook

Non-Streaming Destinations

Destinations that accept events in groups or batches are referred to as non-streaming Destinations:

- Amazon S3 Compatible Stores
- Data Lakes > Amazon S3
- Data Lakes > Amazon Security Lake

- Azure Blob Storage
- Azure Data Explorer when in batching mode
- Exabeam
- Filesystem/NFS
- Google Cloud Storage
- MinIO

The Amazon S3 Compatible Stores Destination can be adapted to send data to downstream services like Databricks and Snowflake, for which Cribl Edge currently has no preconfigured Destination. For details, please contact Cribl Support.

Internal and Special-Purpose Destinations

These special-purpose Destinations route data within your Cribl Edge deployment, or among Workers across distributed or hybrid Cloud deployments:

- Default: Specify a default output from among your configured Destinations.
- Output Router: A "meta-Destination." Configure rules that route data to multiple configured Destinations.
- DevNull: Simply drops events. Preconfigured and active when you install Cribl Edge, so it requires no configuration. Useful for testing.
- Cribl HTTP: Send data among peer Edge Nodes over HTTP.
- Cribl TCP: Send data among peer Edge Nodes over TCP.
- Cribl Stream (Deprecated): Use either Cribl HTTP or Cribl TCP instead.
- **SpaceOut**: This experimental Destination is undocumented. Be careful!

How Does Non-Streaming Delivery Work

Cribl Edge uses a staging directory in the local filesystem to format and write outputted events before sending them to configured Destinations. After a set of conditions is met – typically file size and number of files, further details below – data is compressed and then moved to the final Destination.

An inventory of open, or in-progress, files is kept in the staging directory's root, to avoid having to walk that directory at startup. This can get expensive if staging is also the final directory. At startup, Cribl Edge will check for any leftover files in progress from prior sessions, and will ensure that they're moved to their final Destination. The process of moving to the final Destination is delayed after startup (default delay: 30 seconds). Processing of these files is paced at one file per service period (which defaults to 1 second).

In Cribl.Cloud, using a staging directory is only available on hybrid, customer-managed Edge Nodes.

Batching Conditions

Several **conditions** govern when files are closed and rolled out:

1. File reaches its configured maximum size.
2. File reaches its configured maximum open time.
3. File reaches its configured maximum idle time.

If a new file needs to be open, Cribl Edge will enforce the maximum number of open files, by closing files in the order in which they were opened.

Data Delivery and Persistent Queues

Cribl Stream attempts to deliver data to all Destinations on an at-least-once basis. When a Destination is unreachable, there are three possible behaviors:

- **Block** - Cribl Stream will block incoming events.
- **Drop** - Cribl Stream will drop events addressed to that Destination.
- **Queue** - To prevent data loss, Cribl Stream will write events to a **Persistent Queue** disk buffer, then forward them when a Destination becomes available. (Available on several streaming Destinations.)

For further information about backpressure, see Destination Backpressure Triggers.

You can configure your desired behavior through a Destination's **Backpressure Behavior** drop-down. Where other options are not displayed, Cribl Stream's default behavior is **Block**. For details about all the above behaviors and options, see Persistent Queues.

Configuring Destinations

For each Destination **type**, you can create multiple definitions, depending on your requirements.

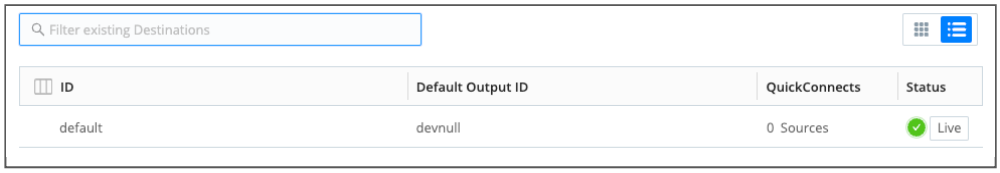
To configure Destinations, from the top nav of a distributed deployment, first click **Manage**, then select a **Fleet** to configure and choose one of the options below. For a single-instance deployment proceed to the options below:

- To access the graphical QuickConnect UI, click **Collect**. Next, select the desired type, and then click either **Add New** or (if displayed) **Select Existing**.

- To access the Routing UI, click **More** > **Destinations**. On the resulting **Manage Destinations** page's tiles, select the desired type, then click **New Destination**.

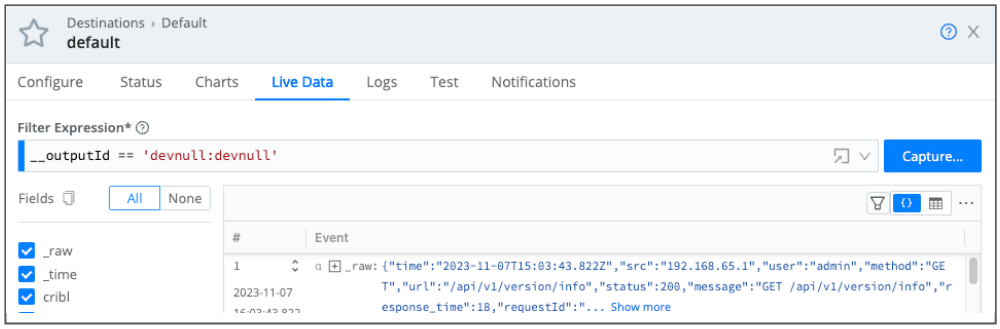
Capturing Outgoing Data

To capture data from a single enabled Destination, you can bypass the Preview pane, and instead capture directly from a **Manage Destinations** page. Just click the **Live** button beside the Destination you want to capture.



Destination > Live button

You can also start an immediate capture from within an enabled Destination's config modal, by clicking the modal's **Live Data** tab.



Destination modal > Live Data tab

Cribl University offers a course titled Sending Data with Edge that provides an illustrated overview. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Cribl Edge courses.

Monitoring Destination Status

Each Destinations's configuration modal offers two tabs for monitoring: **Status** and **Charts**.

Status Tab

The **Status** tab provides details about the Edge Nodes in the Fleet and their status. An icon shows whether the Edge Node is operating normally.

You can click each Edge Node's row to see specific information, for example, to identify issues when the Destination displays an error. The specific set of information provided depends on the Destination type. The data represents only process 0 for each Edge Node.

The content of the **Status** tab is loaded live when you open it and only displayed when all the data is ready. With a lot of busy Edge Nodes in a group, or nodes located far from the Leader, there may be a delay before you see any information.

The statistics presented are reset when the Edge Node restarts.

Charts Tab

The **Charts** tab presents a visualization of the recent activity on the Destination. The following data is available:

- Events in
- Thruput in (events per second)
- Bytes in
- Thruput in (bytes per second)
- Blocked status

This data (in contrast with the status tab) is read almost instantly and does not reset when restarting an Edge Node.

11.1. Internal

11.1.1. Cribl HTTP

The Cribl HTTP Destination sends data to a Cribl HTTP Source in the same Distributed deployment, including Cribl.Cloud, at no charge. It's common to send data from Edge and Stream within the same deployment using a Cribl HTTP Destination and Cribl HTTP Source pair.

The Cribl HTTP Destination is available only in Distributed deployments. In single-instance mode, or for testing, you can substitute it with the Webhook Destination. However, this substitution will not facilitate sending all internal fields, as described below.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

You might choose this Destination over the Cribl TCP Destination in certain circumstances, such as when a firewall or proxy blocks TCP traffic.

How It Works

You can use the Cribl HTTP Destination to transfer data between Workers. If the Cribl HTTP Destination sends data to its Cribl HTTP Source counterpart on another Worker, you're billed for ingress only once – when Cribl first receives the data. All data subsequently relayed to other Workers via a Cribl HTTP Destination/Source pair is not charged.

This use case is common in hybrid Cribl.Cloud deployments, where a customer-managed (on-prem) Node sends data to a Worker in Cribl.Cloud for additional processing and routing to Destinations. However, the Cribl HTTP Destination/Source pair can similarly reduce your metered data ingress in other scenarios, such as on-prem Edge to on-prem Stream.

As one usage example, assume that you want to send data from one Node deployed on-prem, to another that is deployed in Cribl.Cloud. You could do the following:

- Create an on-prem File System Collector (or whatever Collector or Source is suitable) for the data you want to send to Cribl.Cloud.
- Create an on-prem Cribl HTTP Destination.
- Create a Cribl HTTP Source, on the target Stream Worker Group or Edge Fleet in Cribl.Cloud.
- For an on-prem Node configure a File System Collector to send data to the Cribl HTTP Destination, and from there to the Cribl HTTP Source in Cribl.Cloud.

- On Cribl-managed Cribl.Cloud Nodes, make sure that TLS is either disabled on both the Cribl HTTP Destination and the Cribl HTTP Source it's sending data to, or enabled on both. Otherwise, no data will flow. On Cribl.Cloud instances, the Cribl HTTP Source ships with TLS enabled by default.

Configuration Requirements

The key points about configuring this architecture are:

- The Cribl HTTP Destination must be on a Node that is connected to the same Leader as the Cribl HTTP Source(s).
- This Destination's **Cribl endpoint** field must point to the **Address** and **Port** you've configured on its peer Cribl HTTP Source(s).
- Cribl 3.5.4 was a breakpoint in Cribl HTTP Leader/Worker communications. Nodes running the Cribl HTTP Destination on Cribl 3.5.4 and later can receive data only from Nodes running v.3.5.4 and later. Nodes running the Cribl HTTP Destination on Cribl 3.5.3 and earlier can receive data only from Nodes running v.3.5.3 and earlier.

Configuring a Cribl HTTP Destination

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Cribl HTTP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Cribl HTTP**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Load balancing: Set to `No` by default. When toggled to `Yes`, see Load Balancing Settings below. With the default `No` setting, if you notice that Cribl Edge is not sending data to all possible IP addresses, enable **Advanced Settings** > **Round-robin DNS**.

Cribl endpoint: URL of a Cribl Worker to send events to, e.g., `http://localhost:10200`.

The **Cribl endpoint** field appears only when **Load balancing** is toggled to `Off`. Its value must point to the **Address** and **Port** you've configured on the peer Cribl HTTP Source to which you're sending.

Optional Settings

Compression: Codec to use to compress the data before sending. Defaults to `Gzip`.

Backpressure behavior: Specifies whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`. See Persistent Queue Settings below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Load Balancing Settings

Enabling the **Load balancing** toggle displays the following controls.

Exclude Current Host IPs: This toggle determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Cribl Worker Endpoints: In this table, you specify a set of Cribl Workers on which to load-balance data. To specify more Workers on new rows, click **Add Endpoint**. Each row provides the following fields.

- **Cribl Endpoint**: Enter the URL of a Worker to send events to.

  Must point to the **Address** and **Port** configured on a peer Cribl HTTP Source to which you're sending.

- **Load weight**: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.
- The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

Persistent Queue Settings

This section displays when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, Cribl Stream stops queueing and applies the fallback **Queue-full behavior**. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Stream will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear persistent queue: Click this button if you want to flush out files that are currently queued for delivery to this Destination. A confirmation modal will appear. (Appears only after **Output ID** has been defined.)

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Stream Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After`

header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.

- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Stream to cycle through them in the order returned. Only displayed when the General Settings tab's **Load balancing** option is disabled.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Stream will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Exclude fields: Fields to exclude from the event. By default, `__kube_*` and `__metadata` are excluded. This Destination forwards all other Internal Fields.

If you are running Cribl Edge 4.0.x (or earlier) and are using the Kubernetes Metrics Source with this Destination, consider excluding the following fields to reduce event size:

`!__inputId,!__outputId,!__criblMetrics,__* .`

The contents of `__raw` are often redundant with the `_raw` field's contents. Where they are identical, consider excluding one of the two.

Auth Token TTL minutes: The number of minutes before the internally generated authentication token expires, valid values between 1 and 60.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

The following options are added if you enable the General Settings tab's **Load balancing** option:

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that If multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Stream load balancing, IP settings take priority over those from hostnames.)

Internal Fields Loopback to Sources

The Cribl HTTP and Cribl TCP Destinations differ from all other Destinations in the way they handle internal fields: They normally send data back to their respective Cribl Sources – where Cribl internal fields, metrics, and sender-generated fields can all be useful.

These Destinations forward all internal fields by default, except for any that you exclude in **Advanced Settings** > **Exclude fields**.

As examples, if the following fields are present on an event forwarded by a Cribl HTTP or Cribl TCP Destination, they'll be accessible in the ingesting Cribl HTTP/TCP Source: `__criblMetrics`, `__srcIpPort`, `__inputId`, and `__outputId`.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.1.2. Cribl TCP

The Cribl TCP Destination sends data to a Cribl TCP Source in the same Distributed deployment, including Cribl.Cloud, at no charge. It's common to send data from Edge and Stream within the same deployment using a Cribl TCP Destination and Cribl TCP Source pair.

The Cribl TCP Destination is available only in Distributed deployments. In single-instance mode or for testing, you can substitute it with the TCP JSON Destination. However, this substitution will not facilitate sending all internal fields, as described below.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

You might choose this Destination over the Cribl HTTP Destination in certain circumstances, such as when a firewall or proxy allows TCP traffic.

How It Works

You can use the Cribl TCP Destination to transfer data between Workers. If the Cribl TCP Destination sends data to its Cribl TCP Source counterpart on another Worker, you're billed for ingress only once – when Cribl first receives the data. All data subsequently relayed to other Workers via a Cribl TCP Destination/Source pair is not charged.

This use case is common in hybrid Cribl.Cloud deployments, where a customer-managed (on-prem) Node sends data to a Worker in Cribl.Cloud for additional processing and routing to Destinations. However, the Cribl TCP Destination/Source pair can similarly reduce your metered data ingress in other scenarios, such as on-prem Edge to on-prem Stream.

As one usage example, assume that you want to send data from one Node deployed on-prem, to another that is deployed in Cribl.Cloud. You could do the following:

- Create an on-prem File System Collector (or whatever Collector or Source is suitable) for the data you want to send to Cribl.Cloud.
- Create an on-prem Cribl TCP Destination.
- Create a Cribl TCP Source, on the target Stream Worker Group or Edge Fleet in Cribl.Cloud.
- For an on-prem Node configure a File System Collector to send data to the Cribl TCP Destination, and from there to the Cribl TCP Source in Cribl.Cloud.
- On Cribl-managed Cribl.Cloud Nodes, make sure that TLS is either disabled on both the Cribl TCP Destination and the Cribl TCP Source it's sending data to, or enabled on both. Otherwise, no data will flow. On Cribl.Cloud instances, the Cribl TCP Source ships with TLS enabled by default.

Configuration Requirements

The key points about configuring this architecture are:

- The Cribl TCP Destination must be on a Node that is connected to the same Leader as the Cribl TCP Source.
- When you configure the Cribl TCP Destination, its **Address** and **Port** values must point to the **Address** and **Port** you've configured on the Cribl TCP Source.
- Cribl 3.5.4 was a breakpoint in Cribl TCP Leader/Worker communications. Nodes running the Cribl TCP Source on Cribl 3.5.4 and later can send data only to Nodes running v.3.5.4 and later. Nodes running the Cribl TCP Source on Cribl 3.5.3 and earlier can send data only to Nodes running v.3.5.3 and earlier.

Configuring a Cribl TCP Destination

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Cribl TCP**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Cribl TCP**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition.

Load balancing: When toggled to `Yes`, see Load Balancing Settings below.

The following two fields appear **only** with **Load balancing**'s default `No` setting, and must match the **Address** and **Port** you've configured on the peer Cribl TCP Source to which you're sending.

Address: Hostname of the receiver.

Port: Port number to connect to on the host, e.g., `10300`.

Optional Settings

Compression: Codec to use to compress the data before sending. Defaults to `None`.

Throttling: Throttle rate, in bytes per second. Defaults to `0`, meaning no throttling. Multiple-byte units such as KB, MB, GB etc. are also allowed, e.g., `42 MB`. When throttling is engaged, your **Backpressure behavior** selection determines whether Cribl Edge will handle excess data by blocking it, dropping it, or queueing it to disk.

Backpressure behavior: Specifies whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`. See Persistent Queue Settings below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Load Balancing Settings

Enabling the **Load balancing** toggle displays the following controls:

Exclude Current Host IPs

This toggle determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Destinations

The **Destinations** table is where you specify a set of receivers on which to load-balance data.
Click **Add Destination** to specify more receivers on new rows. Each row provides the following fields:

Address: Hostname of a receiver. Optionally, you can paste in a comma-separated list, in `<host>:<port>` format.

Port: Port number to send data to on the host.

Each **Address**/**Port** combination must match the **Address** and **Port** configured on a peer TCP Source to which you're sending.

TLS: Whether to inherit TLS configs from group setting, or disable TLS. Defaults to `inherit`.

TLS servername: Servername to use if establishing a TLS connection. If not specified, defaults to connection host (if not an IP); otherwise, uses the global TLS settings.

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

Persistent Queue Settings

This section displays when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, Cribl Stream stops queueing and applies the fallback **Queue-full behavior**. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Stream will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear persistent queue: Click this button if you want to flush out files that are currently queued for delivery to this Destination. A confirmation modal will appear. (Appears only after **Output ID** has been defined.)

TLS Settings (Client Side)

Use TLS defaults to `No`. When toggled to `Yes`:

Autofill?: This setting is experimental.

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Timeout Settings

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish before retrying. Defaults to `10000`.

Write timeout: Amount of time (in milliseconds) to wait for a write to complete before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Stream Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Auth Token TTL minutes: The number of minutes before the internally generated authentication token expires, valid values between 1 and 60.

Exclude fields: Fields to exclude from the event. By default, `__kube_*` and `__metadata` are excluded. This Destination forwards all other Internal Fields.

If you are running Cribl Edge 4.0.x (or earlier) and are using the Kubernetes Metrics Source with this Destination, consider excluding the following fields to reduce event size:

`!__inputId,!__outputId,!__criblMetrics,__*` .

The contents of `__raw` are often redundant with the `_raw` field's contents. Where they are identical, consider excluding one of the two.

Log failed requests to disk: Toggling to `Yes` makes the payloads of any (future) failed requests available for inspection. See Inspecting Payloads to Troubleshoot Closed Connections below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

The following options are added if you enable the General Settings tab's **Load balancing** option:

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that If multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Stream load balancing, IP settings take priority over those from hostnames.)

Max connections: Constrains the number of concurrent receiver connections, per Worker Process, to limit memory utilization. If set to a number > `0`, then on every DNS resolution period, Cribl Edge will randomly select this subset of discovered IPs to connect to. Cribl Edge will rotate IPs in future resolution periods – monitoring weight and historical data, to ensure fair load balancing of events among IPs.

Inspecting Payloads to Troubleshoot Closed Connections

When a downstream receiver closes connections from this Destination (or just stops responding), inspecting the payloads of the resulting failed requests can help you find the cause. For example:

- Suppose you send an event whose size is larger than the downstream receiver can handle.

- Suppose you send an event that has a `number` field, but the value exceeds the highest number that the downstream receiver can handle.

When **Log failed requests to disk** is enabled, you can inspect the payloads of failed requests. Here is how:

1. In the Destination UI, navigate to the **Logs** tab.
2. Find a log entry with a `connection error` message.
3. Expand the log entry.
4. If the message includes the phrase `See payload file for more info`, note the path in the `file` field on the next line.

Now you have the path to the directory where Cribl Edge is storing payloads from failed requests. At the command line, navigate to that directory and inspect any payloads that you think might be relevant.

Internal Fields Loopback to Sources

The Cribl TCP and Cribl HTTP Destinations differ from all other Destinations in the way they handle internal fields: They normally send data back to their respective Cribl Sources – where Cribl internal fields, metrics, and sender-generated fields can all be useful.

These Destinations forward all internal fields by default, except for any that you exclude in **Advanced Settings** > **Exclude fields**.

As examples, if the following fields are present on an event forwarded by a Cribl HTTP or Cribl TCP Destination, they'll be accessible in the ingesting Cribl HTTP/TCP Source: `__criblMetrics`, `__srcIpPort`, `__inputId`, and `__outputId`.

11.1.3. Disk Spool Destination

The **Disk Spool** Destination writes recent events to disk, so you can spool incoming event data and save it in a predefined file path location. This consistent path can help you find data when you use generic Search datasets.

Type: **Internal** | TLS Support: **N/A** | PQ Support: **No**

Configuring Edge to Spool Event Data

1. From the top nav, click **Manage**, then select the **Fleet** for which you want to add this Destination.
2. Click **Collect**.
3. Click **Add Destination**.
4. Select **Disk Spool** from the resulting drawer's tiles.

The **General Settings** drawer will open.

General Settings

Output ID: Enter a unique name to identify this Disk Spool Destination definition.

Optional Settings

Bucket time span: The amount of time that data is held in each bucket before it's written to disk. The default is 10 minutes (`10m`). Consider the volume of the data source, as high-volume logs can fill a disk up quickly if you enter a long time span.

Data size limit: This is the maximum amount of disk space that spooled data is allowed to consume. Once data reaches this amount, Edge will delete data starting with the oldest buckets.

Data age limit: The duration of time for which Cribl Edge will retain data. Edge will delete data that exceeds the max age entered in this field. Example values are `2h`, `4d`. The default value is 24 hours (`24h`).

Compression: The codec that Edge uses to compress the data. The default is `gzip`.

Partitioning expression: The JavaScript expression that defines how Edge partitions and organizes files into time bucket directories. The default is date-based. If this field is left blank, Cribl Edge will fall back to the

event's `__partition` field value (if present), or otherwise to the root directory of the Output Location and Staging Location.

Use Case Example: When you partition incoming events, Search can find them more efficiently. Imagine you've configured three Sources, each sending event data to the Disk Spool Destination. Enter the partitioning expression `__inputId`. Now you can send data from the three different Sources to individual subdirectories under the spool's time bucket directories.

For help with writing partitioning expressions, check out Improving Partitioning Expressions.

Tags: Add tags that you can use to filter and group Destinations on the Manage Destinations page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Post-Processing

Pipeline: The Pipeline that will process the data before sending it out.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Environment: If you're using GitOps, use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Collect Event Data

Once your Destination is configured, you can begin sending data to it from any Source. We recommend only connecting Sources that do not support disk spooling at the Source level. When you have disk spooling

enabled at the Source *and* you connect the Source to the Disk Spool Destination, event data is duplicated and sent into two different spools.

When you have the Disk Spool Destination configured and Sources connected to it, check that data is flowing using the **Live Data** tab in the Disk Spool Destination.

Next, use Cribl Search to locate spooled data.

Searching for Spooled Data

Disk spooling allows you to store recent events to disk, and makes them available to find with Cribl Search. Typically, data from Edge is sent to and stored directly in a Destination. By caching data in a spool, you can perform data queries in real time without needing to connect to the data's final destination.

For example, let's say you're an IT administrator who is troubleshooting issues with a specific Edge Node. Rather than navigating away from Edge to access the software that's hosting all the Edge data and looking through logs, you can stay in Cribl Edge and run a Search query, returning data for a specific Node over a duration of time you define.

To find spooled data, make sure you have at least one Source connected to the Disk Spool Destination, then open Cribl Search.

1. In the search query box, type `dataset="cribl_edge_spool"` to return all events currently in the spool.
2. Use statements, scope, and processing operators to refine your spooled data. Check out Build a Search for more information and examples.

11.1.4. Cribl Stream (Deprecated)

This Destination was deprecated as of Cribl Edge 3.5, and has been removed as of v.4.0. Please instead use the Cribl TCP or the Cribl HTTP Destination to enable Edge Nodes to send data to peer Nodes.

The Cribl Stream Destination enables Edge Nodes, and/or Cribl Stream instances, to send data to one or multiple Cribl Stream instances.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Use New Destinations Instead

The replacement Destinations, Cribl TCP and Cribl HTTP, don't rely on IP filtering, for the following reasons:

- Load balancers and/or proxies between the Cribl Destination and Cribl Source can change the IP address, resulting in a bad match and rejected ingest.
- A Lookup table of all IP addresses needed to be sent to each Worker Node/Edge Node from the Leader, which is not scalable.
- The Lookup table of IP addresses required constant communication between the Worker Node/Edge Nodes and the Leader, making this fragile and placing an arbitrary reliance on the Leader that shouldn't be there.

Configuring a Cribl Stream Destination

In the **QuickConnect** UI: Click **Add Destination** beside **Destinations**. From the resulting drawer's tiles, select **Cribl Stream**. Next, click **Add Destination** to open a **New Cribl Stream** drawer.

Or, in the **Data Routes** UI: From the top nav of a Cribl Stream instance or Group, select **Data** > **Destinations**. From the top nav of a Cribl Edge instance or Fleet, select **More** > **Destinations**.

From the resulting page's tiles or the **Destinations** left nav, select **Cribl Stream**. Next, click **New Destination** to open a **Cribl Stream** > **New Destination** modal that provides the following options and fields.

General Settings

Output ID: Enter a unique name to identify this Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Address: Hostname of the receiver.

Port: Port number to connect to on the host.

Authentication Settings

Use the **Authentication method** buttons to select one of these options:

- **Manual**: In the resulting **Auth token** field, you can optionally enter an auth token to use in the connection header.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the `authToken` header field value described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

Compression: Codec to use to compress the data before sending. Defaults to `None`.

Throttling: Throttle rate, in bytes per second. Defaults to `0`, meaning no throttling. Multiple-byte units such as KB, MB, GB etc. are also allowed, e.g., `42 MB`. When throttling is engaged, your **Backpressure behavior** selection determines whether Cribl Edge will handle excess data by blocking it, dropping it, or queueing it to disk.

Backpressure behavior: Specifies whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`. See Persistent Queue Settings below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

TLS Settings (Client Side)

Enabled defaults to `No`. When toggled to `Yes`:

Autofill?: This setting is experimental.

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Timeout Settings

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish before retrying. Defaults to `10000`.

Write timeout: Amount of time (in milliseconds) to wait for a write to complete before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.

- `cribl_wp` – Cribl Edge Worker Process that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.1.5. Default

The **Default** Destination simply enables you to specify a default output from among your already-configured Destinations.

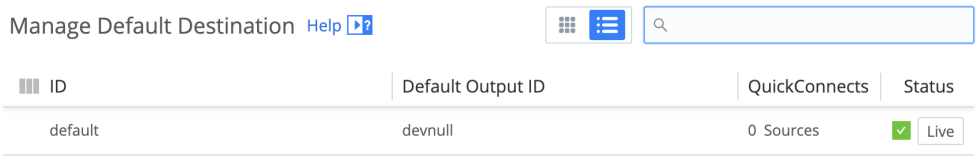
Type: Internal | TLS Support: N/A | PQ Support: N/A

Configuring Cribl Edge's Default Destination

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). A **Default** Destination is preconfigured and ready to use in the right **Destinations** column. Hover over the tile and click its **Configure** button to proceed.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Default**. From the resulting **Manage Default Destination** page, click anywhere on the `default` row to proceed.



Default Destination – click to configure

In the resulting **Destinations** > **Default** drawer or modal, use the **Default Output ID** drop-down to select one of your configured Destinations. After you click **Save**, this will become Cribl Edge's default Destination.

The only other field here is the **Output ID**, whose value is locked to `default`.

Preventing Circular References

If you've configured an Output Router Destination with a branch that points to this Default Destination (`default:default`), you cannot select that Output Router here. This restriction prevents a circular dependency.

11.1.6. DevNull

The DevNull Destination simply drops events. Cribl provides this as a basic output to test Pipelines and Routes.

Type: Internal | TLS Support: N/A | PQ Support: N/A

Configuring Cribl Edge to Forward to DevNull

DevNull requires no configuration: A DevNull Destination is preconfigured and active as soon as you install Cribl Edge.

To verify this, from the top nav, click **Manage**, then select a **Fleet** to configure. Next, select **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Devnull**. Look for the **Live** indicator at the top right.

In the QuickConnect UI, click **Collect** (Edge only), a **DevNull** Destination is preconfigured and ready to use in the right **Destinations** column.

11.1.7. Output Router

Output Routers are meta-destinations that allow for output selection based on rules. Rules are evaluated in order, top-down, with the first match being the winner.

Type: Internal | TLS Support: N/A | PQ Support: N/A

Configuring Cribl Edge to Send to an Output Router

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Output Router**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Output Router**.
Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Router name: Enter a unique name to identify this Router definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Rules: A list of event routing rules. Each provides the following settings:

- **Filter expression**: JavaScript expression to select events to send to output.
- **Output**: Output to send matching events to.
- **Description**: Optionally, enter a description of this rule's purpose.
- **Final**: Toggle to `No` if you want the event to be checked against other rules lower in the stack.

Optional Settings

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Limitations/Options

- An Output Router cannot reference another. This is by design, to avoid circular references.
- Also to avoid circular references, an Output Router cannot reference a Default Destination that points back to Output Router.
- **Events that do not match any of the rules are dropped**. Use a catch-all rule to change this behavior.
- No post-processing (conditioning) can be attached to the Output Router Destination itself. But you are free to use pre-processing Pipelines on the Source tier, and post-processing Pipelines on any or all Destinations that the Output Router references.
- Data can be cloned by toggling the `Final` flag to `No`. (The default is `Yes`, i.e., no cloning.)

Example

Scenario:

- Send all events where `host` starts with `66` to Destination `San Francisco`.
- From the rest of the events: Send all events with `method` field `POST` or `GET` to both `Seattle` and `Los Angeles` (i.e., clone them).
- Send the remaining events to `New York City`.

Router Name: **router66**

Filter Expression	Output	Final
`host.startsWith('66')`	**San Francisco**	Yes
`method=='POST' || method=='GET`	**Seattle**	No
`method=='POST' || method=='GET'`	**Los Angeles**	Yes
`true`	**New York**	Yes

11.2. Amazon

11.2.1. Amazon CloudWatch Logs

Cribl Edge supports sending data to Amazon CloudWatch Logs. Cribl Edge does **not** have to run on AWS in order to deliver data to CloudWatch Logs.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Amazon CloudWatch Logs

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Amazon** > **CloudWatch Logs**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Amazon** > **CloudWatch Logs**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this CloudWatch definition.

Log group name: CloudWatch log group to associate events with.

Log stream prefix: Prefix for CloudWatch log stream name. This prefix will be used to generate a unique log stream name per Cribl Edge instance. (E.g., `myStream_myHost_myOutputId`.)

Region: AWS region where the CloudWatch Logs group is located.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Authentication

Use the **Authentication method** drop-down to select an AWS authentication method.

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The **Secret** option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. A **Create** link is available to store a new, reusable secret.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION` environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for CloudWatch Logs: Toggle to `Yes` to use Assume Role credentials to access CloudWatch Logs.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Endpoint: CloudWatch Logs service endpoint. If empty, defaults to AWS' Region-specific endpoint. Otherwise, use this field to point to a CloudWatchLogs-compatible endpoint.

To proxy outbound HTTP/S requests, see System Proxy Configuration.

Signature version: Signature version to use for signing CloudWatch Logs requests. Defaults to `v4`.

Max queue size: Maximum number of queued batches before blocking. Defaults to `5`.

Max record size (KB, uncompressed): Maximum size of each individual record before compression. For non-compressible data, 1MB (the default) is the maximum recommended size.

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

How the Amazon CloudWatch Logs Destination Batches Events

Amazon CloudWatch Logs specifies that a batch of log events in a single request cannot span more than 24 hours.

When the Amazon CloudWatch Logs Destination transmits a batch of events, it notes the most recent event in the batch (by date-time) and drops all events in the batch that are 24 hours older (or more) than the most recent event.

To monitor dropped events, check **Out of Range Dropped Event Count** under **Status**. Alternately, you can monitor a debug log action to flag how many events are being dropped in a batch.

11.2.2. Amazon Kinesis Streams

Cribl Edge can output events to **Amazon Kinesis Data Streams** records of up to 1MB uncompressed. Cribl Edge does **not** have to run on AWS in order to deliver data to a Kinesis Data Stream.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Meeting the Needs of Different Downstream Receivers

You can use this Destination to send data to different kinds of downstream receivers that have different expectations. In all cases what the Destination sends out are Kinesis Records. This works because the Amazon Kinesis Data Streams Service is flexible about what is actually in a Record, as the two following use cases show.

Sending Multiple Events within One Record

When used to feed data into Amazon Kinesis Data Firehose, this Destination will start with multiple events as NDJSON, gzip-compress them together into a blob, and send that as a single Kinesis Record. This uses the Kinesis PutRecord API call to send the combined events in a single Record.

Firehose will gunzip-uncompress the Record and split the NDJSON out into individual events again. There are also other downstream receivers besides Firehose that can do the same thing.

This approach requires either:

- **Advanced Settings** > **Compression** set to `Gzip`, **or**
- **Advanced Settings** > **Send batched** toggled on.

Meanwhile, if you want Cribl Edge to be the consumer, on the Kinesis Streams Source, set **Record data format** to `Cribl`. See the format details below.

Sending One Event Per Record

In contrast to the first use case, some downstream receivers require a Kinesis Record to contain exactly one Cribl event. Here, this Destination uses the Kinesis PutRecords API call to send multiple events in a single API call, but where each event has a Record to itself. (Note the difference between `PutRecord` in the previous use case and `PutRecords` in this one.)

This approach requires:

- **Advanced Settings** > **Compression** set to `None`, **and**
- **Advanced Settings** > **Send batched** toggled off.

Meanwhile, if you want Cribl Edge to be the consumer, on the Kinesis Streams Source, set **Record data format** to `Newline JSON`. See the format details below.

Configuring Cribl Edge to Output to Amazon Kinesis Data Streams

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Amazon** > **Kinesis**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Amazon** > **Kinesis**.
Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Kinesis definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Stream name: Enter the name of the Kinesis Data Stream to which to send events.

Region: Select the AWS Region where the Kinesis Data Stream is located.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** drops the newest events from being sent out of Cribl Edge, and throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the

throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Authentication

Use the **Authentication Method** drop-down to select an AWS authentication method.

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

For details, see AWS' Actions, resources, and condition keys for Amazon Kinesis Data Streams documentation.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The **Secret** option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. A **Create** link is available to store a new, reusable secret.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION` environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for Kinesis Streams: Toggle to `Yes` to use Assume Role credentials to access Kinesis Streams.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Endpoint: Kinesis Stream service endpoint. If empty, the endpoint will be automatically constructed from the region.

To proxy outbound HTTP/S requests, see System Proxy Configuration.

Signature version: Signature version to use for signing Kinesis stream requests. Defaults to `v4`.

Put request concurrency: Maximum number of ongoing put requests before blocking. Defaults to `5`.

Max record size (KB, uncompressed): Maximum size of each individual record before compression. For non-compressible data, 1MB (the default) is the maximum recommended size.

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Compression: Type of compression to use on records being sent downstream. Defaults to `Gzip`, and when set to `None`, displays the following control:

Send batched: Whether to send multiple events batched as a single NDJSON Kinesis record (the default) or send each event as its own Kinesis record, serialized as JSON. When toggled off, displays the following control:

- **Max records per flush**: Maximum number of records to send in a single request to the Amazon Kinesis Data Streams Service PutRecords API endpoint.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Amazon Kinesis Data Streams Permissions

The following permissions are needed to write to an Amazon Kinesis Stream:

- `kinesis:DescribeStream` - needed in all cases
- `kinesis:PutRecord` - needed only when **Advanced Settings** > **Compression** is set to `Gzip` or **Advanced Settings** > **Send batched** is toggled on.
- `kinesis:PutRecords` - needed only when **Advanced Settings** > **Send batched** is toggled off.

Record Formats in Detail

When sending multiple events within one record, the Record format will be as follows:

- A header line provides information about the payload (the one supported type is shown below).
- Each Kinesis record will contain multiple events serialized as Newline-Delimited JSON (NDJSON).
- Record payloads (including header and body) will be gzip-compressed if **Advanced Settings** > **Compression** is set to `Gzip`.

Sample Kinesis Record

```
{"format":"ndjson","count":8,"size":3960}
{"_raw":"07-03-2018 18:33:51.136 -0700 ERROR TcpOutputFd - Read error. Connection
{"_raw":"07-03-2018 18:33:51.136 -0700 INFO  TcpOutputProc - Connection to 127.0.0
...
```

When sending one event per Record, the Record format will be as follows:

- No header will be included.
- Each Kinesis record will contain a single event serialized as JSON (**not** NDJSON).
- Record payloads will not be compressed.

11.2.3. Amazon S3 Compatible Stores

S3 is a non-streaming Destination type. Cribl Edge does **not** have to run on AWS in order to deliver data to S3.

Stores that are S3-compatible will also work with this Destination type. For example, the S3 Destination can be adapted to send data to services like Databricks and Snowflake, for which Cribl Edge currently has no preconfigured Destination. For these integrations, please contact Cribl Support.

Type: Non-Streaming | TLS Support: Yes | PQ Support: No

Configuring Cribl Edge to Output to S3 Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Amazon** > **S3**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Amazon** > **S3**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this S3 definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

S3 bucket name: Name of the destination S3 Bucket. This value can be a constant, or a JavaScript expression that will be evaluated only at init time. E.g., referencing a Global Variable: `myBucket-${C.vars.myVar}`.

Event-level variables are not available for JavaScript expressions. This is because the bucket name is evaluated only at Destination initialization. If you want to use event-level variables in file paths, Cribl recommends specifying them in the **Partitioning Expression** field (described below), because this is evaluated for each file.

Staging location: Filesystem location in which to locally buffer files before compressing and moving to final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Key prefix: Root directory to prepend to path before uploading. Enter either a constant, or a JS expression (enclosed in single quotes, double quotes, or backticks) that will be evaluated only at init time.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (supported only on Linux, not Windows) exposes a Parquet Settings left tab, where you **must** configure certain options in order to export data in Parquet format.

Optional Settings

Region: Region where the S3 bucket is located.

Partitioning expression: JavaScript expression that defines how files are partitioned and organized. Default is date-based. If blank, Cribl Edge will fall back to the event's `__partition` field value (if present); or otherwise to the root directory of the **Output Location** and **Staging Location**.

Compress: Data compression format used before moving to final destination. Defaults to `gzip` (recommended). This setting is not available when **Data format** is set to `Parquet`.

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

To prevent files from being overwritten, Cribl appends a random sequence of 6 characters to the end of their names. File name prefix and suffix expressions do not bypass this behavior.

For example, if you set the **File name prefix expression** to `CriblExec` and set the **File name suffix expression** to `.csv`, the file name might display as `CriblExec-adPRWM.csv` with `adPRWM` appended.

Backpressure behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication method** drop-down to select one of these options:

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

The values for **Access key** and **Secret key** can be a constant, or a JavaScript expression (such as `${C.env.MY_VAR}`) enclosed in quotes or backticks, which allows configuration with environment variables.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The **Secret** option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. A **Create** link is available to store a new, reusable secret.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION` environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for S3: Toggle to `Yes` to use Assume Role credentials to access S3.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role. This is required only when assuming a role that requires this ID in order to delegate third-party access. For details, see AWS' documentation.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports `c*` wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.
- In Cribl Edge 4.3 and later, the S3 Collector supports ingesting data in the Parquet format. Therefore, data that you export in Parquet format can be replayed via the Collector.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas

> topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported, and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: Set the target memory size for page segments. Generally, set lower values to improve reading speed, or set higher values to improve compression. Value must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files of this size will be closed and moved to final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When this limit is exceeded, on any individual Worker Process, Cribl Edge will close the oldest open files, and move them to the final output location. Defaults to `100`.

Cribl Edge will close files when **any** of the four above conditions is met.

Staging file limit: Maximum number of files that the Destination will allow to wait for upload before it applies backpressure. Defaults to `100`; minimum is `10`; maximum is `4200`. For context, see Troubleshooting below.

Max concurrent file parts: Maximum number of parts to upload in parallel per file. A value of `1` tells the Destination to send the whole file at once. When set to `2` or above, IAM permissions must include those required for multipart uploads. Defaults to `4`; highest allowed value is `10`.

Add Output ID: When set to `Yes` (the default), adds the **Output ID** field's value to the staging location's file path. This ensures that each Destination's logs will write to its own bucket.

For a Destination originally configured in a Cribl Edge version below 2.4.0, the **Add Output ID** behavior will be switched **off** on the backend, regardless of this toggle's state. This is to avoid losing any files pending in the original staging directory, upon Cribl Edge upgrade and restart. To enable this option for such Destinations, Cribl's recommended migration path is:

- Clone the Destination.
- Redirect the Routes referencing the original Destination to instead reference the new, cloned Destination.

This way, the original Destination will process pending files (after an idle timeout), and the new, cloned Destination will process newly arriving events with **Add output ID** enabled.

Remove empty staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Endpoint: S3 service endpoint. If empty, Cribl Edge will automatically construct the endpoint from the AWS Region. To access the AWS S3 endpoints, use the path-style URL format. You don't need to specify the bucket name in the URL, because Cribl Edge will automatically add it to the URL path. For details, see AWS' Path-Style Requests topic.

Object ACL: Object ACL (Access Control List) to assign to uploaded objects.

Storage class: Select a storage class for uploaded objects. Defaults to `Standard`. The other options are: `Reduced Redundancy Storage`; `Standard, Infrequent Access`; `One Zone, Infrequent Access`; `Intelligent Tiering`; `Glacier Flexible Retrieval`; `Glacier Instant Retrieval`; or `Glacier Deep Archive`.

Server-side encryption: Encryption type for uploaded objects – used to enable encryption on data at rest. Defaults to no encryption; the other options are `Amazon S3 Managed Key` or `AWS KMS Managed Key`.

AWS S3 always encrypts data in transit using HTTPS, with default one-way authentication from server to clients. With other S3-compatible stores (such as our native MinIO Destination), use an `https://` URL to invoke in-transit encryption. Two-way authentication is not required to get encryption, and requires clients to possess a certificate.

Signature version: Signature version to use for signing S3 requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Verify if bucket exists: Toggle this to `No` if you can access files within the bucket, but not the bucket itself. This resolves errors of the form: `error initializing...Bucket does not exist`. For context, see Troubleshooting below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

IAM Permissions

The following permissions are always needed to write to an Amazon S3 bucket:

- `s3:ListBucket`
- `s3:GetBucketLocation`
- `s3:PutObject`

If your Destination needs to do multipart uploads to S3, two more permissions are needed:

- `kms:GenerateDataKey`

- `kms:Decrypt`

See the AWS documentation.

Temporary Access via SSO Provider

You can use Okta or SAML to provide access to S3 buckets using temporary security credentials.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

Troubleshooting

Misconfigured or Nonexistent Buckets

If the S3 bucket that the Destination is trying to send files to does not exist or is misconfigured, the Destination logs will reflect the following pattern:

1. Upon initialization, the Destination will log an error that says the bucket cannot be found. This error will not be logged if **Advanced Settings** > **Verify if bucket exists** is turned off.
2. Every time the Destination "rolls" a new file out of staging to try to upload it to S3, the Destination will log an error, until the number of files rolled exceeds the value of **Advanced Settings** > **Max files to stage**.

Once the Destination has begun to apply backpressure:

- The Destination will log a backpressure warning.
- Events will stop flowing through the Destination.
- The Destination will stop creating staging files.
- Although the Destination will periodically attempt to upload the staged files, it will not log any additional failures.

Resources

Cribl University offers an Advanced Troubleshooting > Destination Integrations: S3 short course. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

See also AWS Sources/Destinations & S3-Compatible Stores for information on common errors.

11.2.4. Amazon SQS

Cribl Edge supports sending events to Amazon Simple Queuing Service.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Send Data to Amazon SQS

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Amazon** > **SQS**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Amazon** > **SQS**.
Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this SQS Destination.

Queue name: The name, URL, or ARN of the SQS queue to send events to. This value must be a JavaScript expression (which can evaluate to a constant), enclosed in single quotes, double quotes, or backticks. To specify a non-AWS URL, use the format: `'{url}/<queueName>'`. (E.g., `':port/<myQueueName>'`.)

Queue type: The queue type used (or created). Defaults to `Standard`. `FIFO` (First In, First Out) is the other option.

Optional Settings

Message group ID: This parameter applies only to queues of type FIFO. Enter the tag that specifies that a message belongs to a specific message group. (Messages belonging to the same message group are processed in FIFO order.) Defaults to `cribl`. Use event field `__messageGroupId` to override this value.

Create queue: Specifies whether to create the queue if it does not exist. Defaults to `Yes`.

Region: Region where SQS queue is located.

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by

permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Authentication

Use the **Authentication method** drop-down to select an AWS authentication method.

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The **Secret** option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. A **Create** link is available to store a new, reusable secret.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION`

environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for SQS: Toggle to `Yes` to use Assume Role credentials to access SQS.

AWS account ID: Enter the SQS queue owner's AWS account ID. Leave empty if the SQS queue is in the same AWS account where this Cribl Edge instance is located.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Endpoint: SQS service endpoint. If empty, the endpoint will be automatically constructed from the region.

Signature version: Signature version to use for signing SQS requests. Defaults to `v4`.

Max queue size: Maximum number of queued batches before blocking. Defaults to `100`.

Max record size (KB): Maximum size of each individual record. Per the SQS spec, the maximum allowed value is 256 KB. (the default).

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Max concurrent requests: The maximum number of in-progress API requests before backpressure is applied. Defaults to `10`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

SQS Permissions

The following permissions are needed to write to an SQS queue:

- `sqs:ListQueues`
- `sqs:SendMessage`
- `sqs:SendMessageBatch`
- `sqs:CreateQueue`
- `sqs:GetQueueAttributes`
- `sqs:SetQueueAttributes`
- `sqs:GetQueueUrl`

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields for this Destination:

- `__messageGroupId`
- `__sqsMsgAttrs`
- `__sqsSysAttrs`

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.2.5. Amazon Security Lake

This Destination delivers data to Amazon Security Lake, a fully-managed security data lake service backed by Amazon S3 buckets.

Type: Non-Streaming | TLS Support: Yes | PQ Support: No

Security Lake ingests events described by the Open Cybersecurity Schema Framework (OCSF) and conveyed as Parquet files. This Destination sends the Parquet files in batches. Cribl Edge does not read or replay the data sent to Security Lake.

Cribl recommends that you start with the Amazon Security Lake Integration topic, which will guide you through the following steps in the correct order:

1. Set up Amazon Security Lake.
2. Set up Cribl Stream.
3. Create an Amazon Security Lake custom source, This is an S3 bucket dedicated to a single OCSF event class.
4. Create a Cribl Amazon Security Lake Destination.
5. Connect a Cribl Source to the Amazon Security Lake Destination.
6. Send data to Amazon Security Lake.

As shown above, you'll create and configure this Destination as the fourth of the six overall steps. The Integration topic covers the key points about configuring this Destination. Meanwhile, refer to the present topic as needed for details about any given setting.

Although Cribl Edge does **not** need to be deployed in Cribl.Cloud in order to deliver data to Amazon Security Lake, setup procedures will differ from those documented here. To learn more, please contact Cribl via the `#packs` channel of Cribl Community Slack.

Configuring Cribl Edge to Output to Amazon Security Lake

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Data Lakes** > **Amazon**

Security Lake. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Data Lakes** > **Amazon Security Lake**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

The Integration topic should be your primary guide to setting up Amazon Security Lake. For that reason, some of the instructions below assume that you have noted IDs, values, etc., from procedures you have already completed as directed by the Integration topic.

Output ID: Enter a unique name to identify this Amazon Security Lake Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

S3 bucket name: Name of the destination S3 Bucket you created when setting up Amazon Security Lake. This value can be a constant, or a JavaScript expression that will be evaluated only at init time. E.g., referencing a Global Variable: `myBucket-${C.vars.myVar}`.

Event-level variables are not available for JavaScript expressions. This is because the bucket name is evaluated only at Destination initialization.

Region: Region where the Amazon Security Lake is located.

Account ID: The ID of the AWS Account whose admin enabled Amazon Security Lake.

Custom source name: The name of the custom source you created in Amazon Security Lake to receive data from this Destination.

Staging location: Filesystem location in which to locally buffer files before compressing and moving to final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Optional Settings

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

Backpressure behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

For the **Authentication Method**, you **must** leave the default **Auto** method selected.

This method uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance. The attached IAM role grants Cribl Edge Edge Nodes access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Assume Role

For these settings, you'll enter values that you noted when creating the custom source in Security Lake.

AssumeRole ARN: The custom source's **Provider role ARN**.

External ID: In the custom source, this is the value of the `ExternalId` element of the trust relationship JSON object.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output. Use the drop-down to select the Pack you installed when setting up Cribl Edge.

System fields: Remove any fields specified here.

Parquet Settings

Note that:

- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.

Row group size: Ideal memory size for row group segments. E.g., 128 MB or 1 GB. Affects memory use when writing. Imposes a target, not a strict limit; the final size of a row group may be larger or smaller. Defaults to 16 MB.

Page size: Ideal memory size for page segments. E.g., 1 MB or 128 MB. Generally, lower values improve reading speed, while higher values improve compression. Imposes a target, not a strict limit; the final size of a row group may be larger or smaller. Defaults to 1 MB.

Parquet version: Which version of Parquet your schemas are in determines which data types are supported and how they are represented. Defaults to version 2.6.

Data page version: Serialization format of data pages. Note that not all reader implementations support Data page V2.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files of this size will be closed and moved to final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When this limit is exceeded, on any individual Edge Node Process, Cribl Edge will close the oldest open files, and move them to the final output location. Defaults to `100`.

Cribl Edge will close files when **any** of the four above conditions is met.

Max concurrent file parts: Maximum number of parts to upload in parallel per file. A value of `1` tells the Destination to send the whole file at once. When set to `2` or above, IAM permissions must include those required for multipart uploads. Defaults to `4`; highest allowed value is `10`.

Add Output ID: When set to `Yes` (the default), adds the **Output ID** field's value to the staging location's file path. This ensures that each Destination's logs will write to its own bucket.

Remove staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Endpoint: Security Lake service endpoint. If empty, defaults to AWS' Region-specific endpoint. Otherwise, it must point to Security Lake-compatible endpoint.

Object ACL: Object ACL (Access Control List) to assign to uploaded objects.

Storage class: Select a storage class for uploaded objects. Defaults to `Standard`. The other options are: `Reduced Redundancy Storage`; `Standard, Infrequent Access`; `One Zone, Infrequent Access`; `Intelligent Tiering`; `Glacier`; or `Deep Archive`.

Server-side encryption: Encryption type for uploaded objects – used to enable encryption on data at rest. Defaults to no encryption; the other options are `Amazon S3 Managed Key` or `AWS KMS Managed Key`.

Amazon S3 always encrypts data in transit using HTTPS, with default one-way authentication from server to clients. Two-way authentication is not required to get encryption, and requires clients to possess a certificate.

Signature version: Signature version to use for signing Security Lake requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

IAM Permissions

The following permissions are always needed to write to an Amazon S3 bucket:

- `s3:ListBucket`
- `s3:GetBucketLocation`
- `s3:PutObject`

If your Destination needs to do multipart uploads to S3, two more permissions are needed:

- `kms:GenerateDataKey`
- `kms:Decrypt`

See the AWS documentation.

Creating and Modifying Parquet Schemas

When you need to add a new Parquet schema or modify an existing one:

1. Navigate to the Parquet Schema Library, located at **Processing** > **Knowledge** > **Parquet Schemas**.
2. Create or modify the schema, as described here.

Cribl strongly recommends that to modify an existing schema, you first clone it and give the clone its own distinct name.

3. Return to this Destination, and select the new or modified schema in the **Parquet schema** drop-down.
4. Click **Save**.

Cloning First is Safer Than Just Modifying

Modifying an existing schema in the Schema Library does **not** propagate your modifications to the Destination's copy of the schema.

- Cloning and renaming schemas is the safest approach, because in Step 3 above, it removes any doubt that your Destination will now use the newly-modified version of your schema.
- If you do **not** clone and rename the schema (i.e., you leave the schema name unchanged), you still must re-select the schema in the Destination's **Parquet schema** drop-down, and click **Save**, to bring the modified version into the Destination.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Troubleshooting Resources

Cribl University offers an Advanced Troubleshooting > Destination Integrations: S3 short course. Since Amazon Security Lake relies on S3 buckets, what you learn from the course can augment your Security Lake skills.

To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

11.3. Azure

11.3.1. Azure Blob Storage

Cribl Edge supports sending data to (and replaying specific events from) both Azure Blob Storage, and Azure Data Lake Storage Gen2, which implements a hierarchical namespace over blob data. This Destination can deliver data to Azure whether Cribl Edge is running on Azure, another cloud platform, or on-prem.

Type: Non-Streaming | TLS Support: Yes | PQ Support: No

Configuring Cribl Edge to Output to Azure Blob Storage

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Azure** > **Blob Storage**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Azure** > **Blob Storage**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition.

Container name: Enter the container name. (A container organizes a set of blobs, similar to a directory in a file system.)

Container names can include only lowercase letters, numbers, and/or hyphens (-). This restriction is imposed by Azure.

Blob prefix: Root directory to prepend to path before uploading.

Staging location: Local filesystem location in which to buffer files before compressing and moving them to the final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (supported only on Linux, not Windows) exposes a Parquet Settings left tab, where you **must** configure certain options in order to export data in Parquet format.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Use this default option to enter your Azure Storage connection string directly. Exposes a **Connection string** field for this purpose. (If left blank, Cribl Edge will fall back to `env.AZURE_STORAGE_CONNECTION_STRING`.)
- **Secret**: This option exposes a **Connection string (text secret)** drop-down, in which you can select a stored secret that references an Azure Storage connection string. A **Create** link is available to store a new, reusable secret.

Connection String Format

Either authentication method uses an Azure Storage connection string in this format:

```
DefaultEndpointsProtocol=[http|https];AccountName=<your-account-name>;AccountKey=<y
```

A fictitious example, using Microsoft's recommended HTTPS option, is:

```
DefaultEndpointsProtocol=https;AccountName=storagesample;AccountKey=12345678...32;E
```

Optional Settings

Create container: Toggle to `Yes` to create the configured container in Azure Blob Storage if one does not already exist.

Partitioning expression: JavaScript expression that defines how files are partitioned and organized. Default is date-based. If blank, Cribl Edge will fall back to the event's `__partition` field value (if present); or otherwise to the root directory of the **Output Location** and **Staging Location**.

Compress: Data compression format used before moving to final destination. Defaults to `gzip` (recommended). This setting is not available when **Data format** is set to `Parquet`.

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

To prevent files from being overwritten, Cribl appends a random sequence of 6 characters to the end of their names. File name prefix and suffix expressions do not bypass this behavior.

For example, if you set the **File name prefix expression** to `CriblExec` and set the **File name suffix expression** to `.csv`, the file name might display as `CriblExec-adPRWM.csv` with `adPRWM` appended.

Backpressure behavior: Whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported, and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: Set the target memory size for page segments. Generally, set lower values to improve reading speed, or set higher values to improve compression. Value must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files reaching this size will be closed and moved to the final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Default: `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to final output location. Default: `100`.

Cribl Edge will close files when **either** of the `Max file size (MB)` or the `Max file open time (sec)` conditions are met.

Max concurrent file parts: Maximum number of parts to upload in parallel per file. A value of `1` tells the Destination to send one part at a time – that is, to upload the file's contents sequentially. Defaults to `1`; highest allowed value is `10`.

Add Output ID: When set to `Yes` (the default), adds the **Output ID** field's value to the staging location's file path. This ensures that each Destination's logs will write to its own bucket.

For a Destination originally configured in a Cribl Edge version below 2.4.0, the **Add Output ID** behavior will be switched **off** on the backend, regardless of this toggle's state. This is to avoid losing any files pending in the original staging directory, upon Cribl Edge upgrade and restart. To enable this option for such Destinations, Cribl's recommended migration path is:

- Clone the Destination.
- Redirect the Routes referencing the original Destination to instead reference the new, cloned Destination.

This way, the original Destination will process pending files (after an idle timeout), and the new, cloned Destination will process newly arriving events with **Add output ID** enabled.

Remove staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable

interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.3.2. Azure Data Explorer

Cribl Edge supports sending data to the Azure Data Explorer (ADX) managed data analytics service; you can then run Kusto queries against the data. This Destination can deliver data to Azure whether Cribl Edge is running on Azure, another cloud platform, or on-prem.

ADX stores log data in databases, which in turn contain one or more of the tables defined in the Azure namespace. You must create a separate Cribl Edge ADX Destination for each table that will store your data.

Type: Non-Streaming **or** Streaming (configurable) | TLS Support: Yes | PQ Support: No

Prerequisites

Before you configure the Destination, you need to collect some information from your Azure deployment. Start from **Home** (`portal.azure.com`), then navigate to the locations described below and collect the items specified.

Information About Your ADX Cluster

From **Home**, navigate to:

Azure Data Explorer Clusters > `your_Azure_cluster_name` > **Overview** > **Essentials**

- Copy the **URI**. In Cribl Edge, this will be the value for **General Settings** > **Cluster base URI**. Make a copy of the URI with `/.default` appended – this will be the value of **Authentication Settings** > **Scope**.
- Copy the **Data Ingestion URI**. In Cribl Edge, this will be the value for **General Settings** > **Ingestion service URI**.

Information About Your Azure Database

From **Home**, navigate to:

Azure Data Explorer Clusters > `your_Azure_cluster_name` | **Databases** > `database_name` | **Query**

- Note the name of your desired target table – this table name will be the value of **General Settings** > **Table name**.

Information About Your Azure App

First, from **Home**, navigate to:

App registrations > `your_Azure_app_name` | **Overview** | **Essentials**

- Copy the **Application (client) ID** – this will be the value of **Authentication Settings** > **Client ID**.
- Copy the **Directory (tenant) ID** – this will be the value of **Authentication Settings** > **Tenant ID**.

Nex, from **Home**, navigate to:

App registrations > `your_Azure_app_name` | **Certificates & secrets**

Now what you do depends on which method for **Authentication** you plan to use in Cribl Edge.

To use the Client secret method

- In the **Client secrets** tab, click **New client secret** and copy the **Value** of the new secret. In Cribl Edge, this will be the value of **Client secret**.

To use the Client secret (text secret) method

- In the **Client secrets** tab, click **New client secret** and copy the **Value** of the new secret. In Cribl Edge, this will be the value of **Create new secret** > **Value**.

To use the Certificate method

After you have generated a certificate and keys (if you do this later on in the configuration process, return to this section then):

- In the **Certificates** tab, click **Upload Certificate**. In the resulting drawer, follow the prompts to upload your certificate, and click **Add** to close the drawer.

The chosen certificate will then appear in the **Certificates** tab's list.

Data Mapping Prerequisites

Finally, you might need to note or copy data mapping information to obtain values for some of the **General Settings**.

ADX uses a data mapping to map fields from incoming data to fields in the target table. To see the data mapping, you'll run a Kusto query against your target table.

Here's the query syntax:

```
.show table EntityName ingestion MappingKind mapping MappingName
```

Here's an example query:

```
.show table SyslogTable ingestion json mapping "SyslogMapping"
```

The Kusto query that originally created this mapping would look like this:

```
.create table SyslogTable ingestion json mapping "SyslogMapping"
```

In Cribl Edge, you can use the names of mappings like the above example in the **General Settings** > **Data mapping** field. Alternatively, you can create a mapping without naming it, then copy and paste it (as a JSON object) into the Cribl Edge UI. To do this, toggle **General Settings** > **Mapping object** on. This hides the default **Data mapping** text field and opens a text window that expects JSON.

Configuring Cribl Edge to Output to Azure Data Explorer

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Azure** > **Azure Data Explorer**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Azure** > **Azure Data Explorer**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

This Destination supports two **ingestion modes**:

- **Batching** (the default) stages data in a storage container from which ADX pulls batches of data, then writes them to the target ADX table. Batching works well when the Destination needs to ingest large amounts of data in short periods of time.
- **Streaming** sends data as HTTP request payloads, directly to the target ADX table, without staging them in a container first. Streaming can achieve lower latency than batching, as long as the data that the Destination ingests arrives in relatively small amounts. When in this mode, the Retries section (left tab) becomes available.

Output ID: Enter a unique name to identify this Destination definition.

Ingestion mode: Use the buttons to select **Batching** mode (the default) or **Streaming** mode.

Cluster base URI: Enter the base URI for your ADX cluster.

Ingestion service URI: Displayed only in **Batching** mode. Enter the Ingestion Service URI for your ADX cluster.

Database name: Enter the name of the ADX database where the target table resides.

Table name: Enter the name of the target table.

Validate database settings: When you save or start the Destination, validates the database name and credentials that you have entered in these settings. Also validates the table name, except when **Add mapping object** (below) is on. Defaults to `Yes`. Disable if your Azure app does not have **both** the **Database Viewer** and the **Table Viewer** role.

Add mapping object: Displayed only in **Batching** mode, this control is toggled off by default. Toggle on when you want to paste in a data mapping as a JSON object, instead of using a named data mapping. See Data Mapping Prerequisites above.

Data mapping: If **Add mapping object** is off, enter the name of the desired data mapping. If **Add mapping object** is on, enter the desired data mapping as a JSON object.

Compress: By default, this Destination compresses data using `gzip`; otherwise it sends the data uncompressed. This setting is not available when **Data format** is set to `Parquet`.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (available only in **Batching** mode, and supported only on Linux, not Windows) exposes a **Parquet Settings** tab, where you select the Parquet schema.

Backpressure behavior: Whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`. When **Ingestion mode** is set to **Streaming** mode, the **Persistent Queue** option becomes available. See **Persistent Queue Settings** below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication Settings

Microsoft Entra ID authentication endpoint: From the drop-down, select the Microsoft Entra ID endpoint that provides authentication tokens. To understand the available choices, see the Microsoft Entra documentation.

Tenant ID: Directory ID (tenant identifier) in Microsoft Entra ID.

Client ID: `client_id` to pass in the OAuth request parameter.

Scope: Scope to pass in the OAuth request parameter.

Use the **Authentication method** buttons to select one of these options:

- **Client secret**: Use this default option to enter the client secret that you generated for your app in the Azure portal.
- **Client secret (text secret)**: Use this option to select (or create) a stored text secret.
- **Certificate**: Use this option to select the certificate whose public key you register (or will register) for your app in the Azure portal. Or, create a new one.

Persistent Queue Settings

This tab is displayed when **General Settings** > **Ingestion mode** is set to **Streaming** mode and **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Edge Nodes (with an Enterprise plan), this tab should expose only the destructive **Clear Persistent Queue** button (described below in this section).

Due to Known Issue CRIBL-21335, the on-prem Persistent Queue settings are still displayed in Cribl.Cloud – but you should ignore all controls other than the **Clear Persistent Queue** button.

The other controls are not needed because Cribl.Cloud automatically allocates a maximum queue size of 1 GB disk space per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the `Block` option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: The target memory size for page segments. Generally, lower values improve reading speed, while higher values improve compression. Must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Retries

This tab is displayed when **General Settings** > **Ingestion mode** is set to **Streaming** mode.

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, `429 (Too Many Requests)` is the only response code configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.

- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

The controls on this tab vary depending on the **Ingestion mode** selected on the **General Settings** tab. Below you'll find one dedicated section for the Batching, and another for the Streaming controls.

Only the **Environment** control, at the bottom of the tab, is common to both:

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Advanced Settings for Batching Mode

Flush immediately: Toggle on to bypass the data management service's aggregation mechanism. See the ADX documentation.

Retain blob on success: By default, Cribl Edge deletes each data blob once ingestion is complete. Toggle this on if you prefer to retain blobs instead.

Extent tags: Optionally, add strings or tags to partitions of the target table – Azure calls these extents or data shards.

Enforce uniqueness via tag values: Use the **Add value** button to specify a list of `ingest-by` values. Cribl Edge will then check the `ingest-by` tags of incoming extents and discard those whose values match a listed value. This mechanism for avoiding ingesting duplicate extents uses the `ingestIfNotExists` property, as described in the ADX documentation.

Report level: Level for ingestion status reporting. Options are `DoNotReport`, `FailuresOnly` (the default), and `FailuresAndSuccesses`.

Report method: Target for ingestion status reporting. Options are `Queue` (the default), `Table`, and `QueueAndTable`.

Additional fields: Optionally, enter additional configuration properties to send to the ingestion service.

Staging location: Local filesystem location in which to buffer files before compressing and moving them to the final destination. Cribl recommends that this location be stable and high-performance. Defaults to `/tmp`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks.

Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

To prevent files from being overwritten, Cribl appends a random sequence of 6 characters to the end of their names. File name prefix and suffix expressions do not bypass this behavior.

For example, if you set the **File name prefix expression** to `CriblExec` and set the **File name suffix expression** to `.csv`, the file name might display as `CriblExec-adPRWM.csv` with `adPRWM` appended.

Max file size (MB): Maximum uncompressed output file size. Files reaching this size will be closed and moved to the storage container. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to the storage container. Defaults to `300` seconds (5 minutes).

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to the storage container. Default: `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to the storage container. Default: `100`.

Max concurrent file parts: Maximum number of parts to upload in parallel per file. A value of `1` tells the Destination to send one part at a time – that is, to upload the file's contents sequentially. Defaults to `1`; highest allowed value is `10`.

Add output ID: Toggle on if you want Cribl Edge to append the Destination name to staging directory pathnames. This can make it easier to organize and troubleshoot when you have multiple Destinations

populating staging directories.

Remove empty staging dirs: When toggled on (the default), Cribl Edge deletes empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Advanced Settings for Streaming Mode

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Validate server certs: When toggled on (the default) Cribl Edge will reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (such as the system's CA, for example).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Keep alive: By default, Cribl Edge sends `Keep-Alive` headers to the remote server and preserves the connection from the client side up to a maximum of 120 seconds. Toggle this off if you want Cribl Edge to close the connection immediately after sending a request.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.3.3. Azure Event Hubs

Cribl Edge supports sending data to Azure Event Hubs.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Azure Event Hubs uses a binary protocol over TCP. It does not support HTTP proxies, so Cribl Edge must send events directly to receivers. You might need to adjust your firewall rules to allow this traffic.

See Microsoft's Compare Azure Event Hubs Tiers topic to configure tiers whose features and quotas match your desired data volume and throughput.

Configuring Cribl Edge to Output to Azure Event Hubs

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Azure** > **Events Hub**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Azure** > **Events Hub**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Azure Event Hubs definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Brokers: List of Event Hub Kafka brokers to connect to. (E.g., `yourdomain.servicebus.windows.net:9093`.) Find the hostname in Shared Access Policies, in the host portion of the primary or secondary connection string.

Event Hub name: The name of the Event Hub (a.k.a., Kafka Topic) on which to publish events. Can be overwritten using the `__topicOut` field.

Optional Settings

Acknowledgments: Control the number of required acknowledgments. Defaults to `Leader`. Setting this control to `Leader` or `All` may significantly slow throughput. If the impact is too great for your use case, set this control to `None`.

Record data format: Format to use to serialize events before writing to the Event Hub Kafka brokers. Defaults to `JSON`.

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

TLS Settings (Client Side)

Enabled Defaults to `Yes`.

Validate server certs: Defaults to `Yes`.

Authentication

Authentication parameters to use when connecting to brokers. Using TLS is highly recommended.

Enabled: With the default `Yes` setting, this section's remaining settings are displayed, and all are required settings.

SASL mechanism: SASL (Simple Authentication and Security Layer) authentication mechanism to use with Kafka brokers. Defaults to `PLAIN`, which exposes Basic Authentication options that rely on Azure Event Hubs connection strings. Select `OAUTHBEARER` to enable OAuth Authentication via a different set of options.

Basic Authentication

Selecting the `PLAIN` SASL mechanism provides the options listed in this section.

Username: The username for authentication. For Event Hubs, this should always be `$ConnectionString`.

Authentication method: Use the buttons to select one of these options:

- **Manual**: Use this default option to enter your Event Hubs connection string's primary or secondary key from the Event Hubs workspace. Exposes a **Password** field for this purpose.
- **Secret**: This option exposes a **Password (text secret)** drop-down, in which you can select a stored secret that references an Event Hubs connection string. The secret can reside in Cribl Edge's internal secrets manager or (if enabled) in an external KMS. A **Create** link is available if you need a new secret.

Connection String Format

Either authentication method above uses an Azure Event Hubs connection string in this format:

```
Endpoint=sb://<FQDN>/;SharedAccessKeyName=<your-shared-access-key-name>;SharedAccessKey=<your-shared-access-key-value>
```

A fictitious example is:

```
Endpoint=sb://dummynamespace.servicebus.windows.net/;SharedAccessKeyName=DummyAccessKe
```

OAuth Authentication

Selecting the `OAUTHBEARER` SASL mechanism provides the options listed in this section.

Microsoft Entra ID authentication endpoint: Specifies the Microsoft Entra ID endpoint from which to acquire authentication tokens. Defaults to `https://login.microsoftonline.com`. You can instead select `https://login.microsoftonline.us` or `https://login.partner.microsoftonline.cn`.

Client ID: Enter the `client_id` to pass in the OAuth request parameter.

Tenant identifier: Enter your Azure Active Directory subscription's directory ID (tenant ID).

Scope: Enter the scope to pass in the OAuth request parameter. This will be of the form: `https://<Event-Hubs-Namespace-Host-name>/.default`. (E.g., for an Event Hubs Namespace > Host name: `goatyoga.servicebus.windows.net`, **Scope**: `https://goatyoga.servicebus.windows.net/.default`.)

Authentication method: Use the buttons to select one of these options:

- **Manual**: This default option exposes a **Client secret** field, in which to directly enter the `client_secret` to pass in the OAuth request parameter.
- **Secret**: Exposes a **Client secret (text secret)** drop-down, in which you can select a stored secret that references the `client_secret`. A **Create** link is available to define a new secret.

- **Certificate**: Exposes a **Certificate name** drop-down, in which you can select a stored certificate. A **Create** link is available to define a new cert.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (KB, uncompressed): Maximum size (KB) of each record batch before compression. Setting should be < `message.max.bytes` settings in Kafka brokers. Defaults to `768`.

Max events per batch: Maximum number of events in a batch before forcing a flush. Defaults to `1000`.

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Connection timeout (ms): Maximum time to wait for a successful connection. Defaults to `10000` ms, i.e., 10 seconds. Valid range is `1000` to `3600000` ms, i.e., 1 second to 1 hour.

Request timeout (ms): Maximum time to wait for a successful request. Defaults to `60000` ms, i.e., 1 minute.

Max retries: Maximum number of times to retry a failed request before the message fails. Defaults to `5`; enter `0` to not retry at all.

Authentication timeout (ms): Maximum time to wait for Kafka to respond to an authentication request. Defaults to `1000` (1 second).

Reauthentication threshold (ms): If the broker requires periodic reauthentication, this setting defines how long before the reauthentication timeout Cribl Edge initiates the reauthentication. Defaults to `10000` (10 seconds).

A small value for this setting, combined with high network latency, might prevent the Destination from reauthenticating before the Kafka broker closes the connection.

A large value might cause the Destination to send reauthentication messages too soon, wasting bandwidth.

The Kafka setting `connections.max.reauth.ms` controls the reuthentication threshold on the Kafka side.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Working with Event Timestamps

By default, when an incoming event reaches this Destination, the underlying Kafka JS library adds a `timestamp` field set to the current time at that moment. The library gets the current time from the `Date.now()` JavaScript method. Events that this Destination sends to downstream Kafka receivers include this `timestamp` field.

Some production situations require the `timestamp` value to be the time an incoming event was originally created by an upstream sender, not the current time of its arrival at this Destination. For example, suppose the events were originally created over a wide time range, and arrive at the Destination hours or days later. In this case, the original creation time might be important to retain in events Cribl Edge sends to downstream Kafka receivers.

Here is how to satisfy such a requirement:

1. In a Pipeline that routes events to this Destination, use an Eval Function to add a field named `__kafkaTime` and write the incoming event's original creation time into that field. The value of `__kafkaTime` **must** be in UNIX epoch time (either seconds or milliseconds since the UNIX epoch – both will work). For context, see this explanation of the related fields `_time` and `__origTime`.
2. This Destination will recognize the `__kafkaTime` field and write its value into the `timestamp` field. (This is similar to the way you can use the `__topicOut` field to overwrite a topic setting, as described above.)

If the `__kafkaTime` field is **not** present, the Destination will apply the default behavior (using `Date.now()`) described previously.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__topicOut`
- `__key`
- `__headers`
- `__keySchemaIdOut`
- `__valueSchemaIdOut`

Troubleshooting Resources

Cribl University offers an Advanced Troubleshooting > Destination Integrations: Azure Event Hubs short course. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

11.3.4. Azure Monitor Logs

Cribl Edge supports sending data to Azure Monitor Logs.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Azure Monitor Logs

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Azure** > **Monitor Logs**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Azure** > **Monitor Logs**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Azure Monitor Logs definition.

Log type: The Record Type of events sent to this LogAnalytics workspace. Defaults to `Cribl`. Use only letters, numbers, and `_` characters. (See Microsoft's Azure Monitor documentation.) Can be overwritten by an event's `__logType` field.

Authentication Settings

Authentication method: Use the buttons to select one of these options:

- **Manual**: Displays fields in which to enter your Azure Log Analytics **Workspace ID** and your Primary or Secondary Shared **Workspace key**. See the Azure Monitor documentation.
- **Secret**: This option exposes a **Secret key pair** drop-down, in which you can select a stored secret that references the credentials described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

DNS name of API endpoint: Enter the DNS name of the Log API endpoint that sends log data to a Log Analytics workspace in Azure Monitor. Defaults to: `.ods.opinsights.azure.com.` Cribl Edge will add a prefix and suffix around this DNS name, to construct a URI in this format: `https://<Workspace_ID><your_DNS_name>/api/logs?api-version=<API version>`.

Resource ID: Resource ID of the Azure resource to associate the data with. This populates the `_ResourceId` property, and allows the data to be included in resource-centric queries. (Optional, but if this field is not specified, the data will not be included in resource-centric queries.)

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).

- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1024` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low settings could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Azure Monitor Limitations

The Azure Monitor Logs architecture limits the number of columns per table, characters per column name, and other parameters. For details, see Microsoft's Azure Monitor Service Limits topic.

Azure will drop logs if your data exceeds these limits. To diagnose this, you can search in the Azure Data Explorer console with a query like this:

```
Operation | summarize count() by Detail
```

...for error messages of this form:

```
Data of type <type> was dropped: The number of custom fields <number> is above the limit of 500 fields per data type.
```

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular Destination when there is a constant flow of events.
- If keepalives are not supported by the server (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between destination cluster nodes.

11.3.5. Azure Sentinel

Cribl Edge can send log and metric events to the Azure Sentinel SIEM. This Destination encrypts and sends events via the HTTPS protocol. (These docs use the terms “Azure Sentinel” and “Microsoft Sentinel” interchangeably.)

Type: Streaming | TLS Support: No | PQ Support: Yes

Prerequisites

Before configuring your Azure Sentinel Destination(s), you have to prepare the Azure workspace, create data collection rules (DCRs), and obtain the ingestion URL that will receive the data from your Destination(s).

Complete these preparatory steps, which are described in Azure Sentinel Integration, before configuring the Azure Sentinel Destination. That topic explains the overall Sentinel/Cribl workflow, provides necessary context, and shows you how to obtain the values you’ll need to configure the Destination.

Behind the scenes, this Destination uses the Azure Monitor Logs Ingestion API.

Configuring Cribl Edge to Output to Azure Sentinel

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer’s tiles, select **Azure Sentinel**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page’s tiles or the **Destinations** left nav, select **Azure Sentinel**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

If you intend to send multiple event types, or tables, to Sentinel, you’ll need a separate Destination for each event type. In this case, Cribl recommends you use a single URL to ingest all the data, with an Output Router to route each event type to the correct Azure Sentinel Destination.

General Settings

Output ID: Enter a unique name to identify this HTTP output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Endpoint configuration: Method for configuring the endpoint. Options include:

- **URL** – lets you directly enter the data collection endpoint. This method is the simplest way of configuring the endpoint. See Obtaining a URL for more information.
- **ID** – lets you enter individual IDs that Cribl Edge uses to create the URL used as the data collection endpoint.

Selecting `URL` exposes the following field.

URL: Endpoint URL to send events to. The internal field `__url`, where present in events, will override the `URL` and `ID` values. See Internal Fields below.

Selecting `ID` exposes the following fields.

Data collection endpoint: Data collection endpoint (DCE) in the format `https://<endpoint-name>.<identifier>.<region>.ingest.monitor.azure.com`.

Data collection rule ID: Immutable ID for the data collection rule (DCR).

Stream name: Name of the Sentinel table in which to store events.

Optional Settings

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication Settings

- **Login URL**: Endpoint for the OAuth API call, which is expected to be a POST call. This URL takes the form `https://login.microsoftonline.com/<tenant-id>/oauth2/v2.0/token`.
- **OAuth secret**: Secret parameter value to pass in the API call's request body.
- **Client ID**: JavaScript expression used to compute the application (client) ID for the Azure application. Can be a constant.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are not authorized by a trusted CA (e.g., the system's CA). Toggle this to `No` if you want Cribl Edge to accept such certificates. Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: By default, Cribl Edge enables gzip-compress to compress the payload body before sending. Toggle this off if you want Cribl Edge not to gzip-compress the payload body.

Keep alive: By default, Cribl Edge sends `Keep-Alive` headers to the remote server and preserves the connection from the client side up to a maximum of 120 seconds. Toggle this off if you want Cribl Edge to close the connection immediately after sending a request.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1000` KB. Be aware that high values can cause high memory usage per Worker Node, especially if a dynamically constructed URL (see Internal Fields) causes this Destination to send events to more than one URL. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass to all events as additional HTTP headers. Values will be sent encrypted. You can also add headers dynamically on a per-event basis in the `__headers` field. See Internal Fields below.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__criblMetrics`
- `__url`
- `__headers`

If an event contains the internal field `__criblMetrics`, Cribl Edge will send it to the HTTP endpoint as a metric event. Otherwise, Cribl Edge will send it as a log event.

If an event contains the internal field `__url`, that field's value will override the **General Settings** > **URL** value. This way, you can determine the endpoint URL dynamically.

For example, you could create a Pipeline containing an Eval Function that adds the `__url` field, and connect that Pipeline to your Azure Sentinel Destination. Configure the Eval Function to set `__url`'s value to a URL that varies depending on a global variable, or on some property of the event, or on some other dynamically-generated value that meets your needs.

If an event contains the internal field `__headers`, that field's value will be a JSON object containing Name-value pairs, each of which defines a header. By creating Pipelines that set the value of `__headers` according to conditions that you specify, you can add headers dynamically.

For example, you could create a Pipeline containing Eval Functions that add the `__headers` field, and connect that Pipeline to your Azure Sentinel Destination. Configure the Eval Functions to set `__headers` values to Name-value pairs that vary depending on some properties of the event, or on dynamically-generated values that meet your needs.

Here's an overview of how to add headers:

- To add "dynamic" (a.k.a. "custom") headers to some events but not others, use the `__headers` field.
- To define headers to add to **all** events, use **Advanced Settings** > **Extra HTTP Headers**.
- An event can include headers added by both methods. So if you use `__headers` to add `{ "api-key": "foo" }` and **Extra HTTP Headers** to add `{ "goat": "Kid A" }`, you'll get both headers.
- Headers added via the `__headers` field take precedence. So if you use `__headers` to add `{ "api-key": "foo" }` and **Extra HTTP Headers** to add `{ "api-key": "bar" }`, you'll get only one header, and that will be `{ "api-key": "foo" }`.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, it will throw away the connection and attempt a new one. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), Cribl Edge will establish a new connection for the next request.

- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable **Round-robin DNS** to better balance distribution of events among destination cluster nodes.

11.4. Elastic

11.4.1. Elasticsearch

Cribl Edge can send events to an Elasticsearch cluster using the Bulk API. As of v.3.3, Cribl Edge supports Elastic data streams.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output to Elasticsearch

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click **Routing** > **QuickConnect** (Stream) or **Collect** (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Elasticsearch**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Elasticsearch**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Elasticsearch Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Load balancing: When enabled (default), lets you specify multiple bulk API URLs and load weights. With the `No` setting, if you notice that Cribl Edge is not sending data to all possible IP addresses, enable **Advanced Settings** > **Round-robin DNS**.

Bulk API URL or Cloud ID: Specify either an Elasticsearch cluster or Elastic Cloud deployment to send events to. For an Elasticsearch cluster, enter a URL (e.g., `http://<myElasticCluster>:9200/_bulk`). For an Elastic Cloud deployment, enter its Cloud ID. This setting is not available when **Load balancing** is enabled.

Bulk API URLs

Use the **Bulk API URLs** table to specify a known set of receivers on which to load-balance data. To specify more receivers on new rows, click **Add URL**. Each row provides the following fields:

URL: Specify the URL to an Elastic node to send events to – e.g., `http://elastic:9200/_bulk`

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

When you first enable load balancing, or if you edit the load weight once your data is load–balanced, give the logic time to settle. The changes might take a few seconds to register.

Index or data stream: Enter a JavaScript expression that evaluates to the name of the Elastic data stream or Elastic index where you want events to go. The expression is evaluated for each event, can evaluate to a constant value, and must be enclosed in quotes or backticks. An event's `__index` field can overwrite the index or data stream name.

Authentication Settings

Authentication enabled: When toggled to `Yes`, use the **Authentication method** buttons to select one of these options:

Manual: Enter your credentials directly in the resulting **Username** and **Password** fields.

Secret: Exposes a **Credentials secret** drop-down, in which you can select a stored secret that references the credentials described above. A **Create** link is available to store a new, reusable secret.

Manual API Key: Exposes an **API key** field to directly enter your Elasticsearch API key.

Secret API Key: Exposes an **API key (text secret)** drop-down, in which you can select a stored text secret that references your Elasticsearch API key. A **Create** link is available to store a new, reusable secret.

Optional Settings

Exclude current host IPs: This toggle appears when **Load balancing** is set to `Yes`. It determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Type: Specify document type to use for events. An event's `__type` field can overwrite this value.

Backpressure behavior: Specify whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. Cribl Edge will append `/<worker-id>/<output-id>` to this field's value.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the `Block` option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by

permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are **not** authorized by a trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned. (This option is visible only when the **General Settings** > **Load balancing** option is set to `No`.)

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (s): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Extra parameters: Name-value pairs to pass as additional parameters. If you are using Elastic ingest pipelines, specify an extra parameter whose name is `pipeline` and whose value is the name of your pipeline, similar to these examples.

Elastic version: Determines how to format events. For Elastic Cloud, you must explicitly set version `7.x`. For other Elasticsearch clusters, the `Auto` default will discover the downstream Elasticsearch version automatically, but you have the option to explicitly set version `6.x` or `7.x`.

Elastic pipeline: To send data to an Elastic Ingest pipeline, optionally enter that pipeline's name as a constant. Or, enter a JavaScript expression that evaluates outgoing events and sends matching events to the desired Elastic Ingest pipeline(s). Cribl Edge will first interpret your entry as a constant and try to find a matching value in the event. If it finds no matching value, it will evaluate your **Elastic pipeline** entry as an expression.

For example, the expression `sourcetype=='access_common'?'cribl_pipeline':undefined` matches events whose `sourcetype` is `access_common`, and sends them to an Elastic Ingest pipeline named `cribl_pipeline`.

You can also specify the name of the pipeline in an event field. For example, `myPipelineField` would use the value from the event's `myPipelineField` property (if present) to identify the Elastic Ingest pipeline to process the event. Alternately, the expression `myPipelineField != null ? myPipelineField : undefined` would also identify this field. If the event does not contain such a field, `myPipelineField != null ? myPipelineField : 'theDefaultIndex'` will provide a default index. With this approach, you can override the Elastic Ingest pipeline at the event level.

See also the Elasticsearch Source documentation.

The next two fields appear only when the **General Settings** > **Load balancing** option is set to `Yes`.

DNS resolution period (seconds): Re-resolve any hostnames each time this interval recurs, and pick up destinations from the A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds.

Include document_id: Toggle this setting to `No` to omit the `document_id` field when sending events to an Elastic TSDS (time series data stream).

Write action: Action to use when writing events. Set this option to `Create` (default) when writing to a data stream, which is append-only. Set this option to `Index` only when you write directly to an index and need to update existing records. `Index` will fail if you use it to write to a data stream.

Failed request logging mode: Determines which data is logged when a request fails. Use the drop-down to select one of these options:

- `None` (default).
- `Payload`.
- `Payload + Headers`. Use the **Safe Headers** field below to specify the headers to log. If you leave that field empty, all headers are redacted, even with this setting.

Safe headers: List the headers you want to log, in plain text.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Field Normalization

This Destination normalizes the following fields:

- `_time` becomes `@timestamp` at millisecond resolultion.

- `host.name` is set to `host`.

See also our Elasticsearch Source documentation's **Field Normalization** section.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__id`
- `__type`
- `__index`
- `__host`

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname with load balancing disabled, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between Elasticsearch cluster nodes.

11.4.2. Elastic Cloud

Cribl Edge can send events to Elastic Cloud.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Elastic Cloud

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

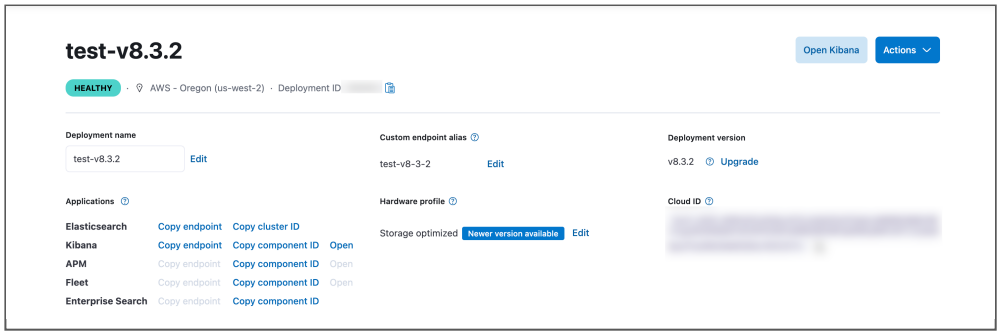
To configure via the graphical QuickConnect UI, click **Routing** > **QuickConnect** (Stream) or **Collect** (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Elastic Cloud**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Elastic Cloud**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Elastic Cloud Destination definition.

Cloud ID: Enter the Cloud ID of the Elastic Cloud environment where you want to send events. You'll find it on the Deployments overview page of your Elastic Cloud environment.



Getting your Elastic Cloud ID

Data stream or index: Enter a JavaScript expression that evaluates to the name of the Elastic data stream or Elastic index where you want events to go. The expression is evaluated for each event, can evaluate to a constant value, and must be enclosed in quotes or backticks. An event's `__index` field can overwrite the index or data stream name.

Elastic pipeline: Enter the name of the Elastic ingest pipeline (optional). You can use an expression to override the pipeline attribute received from the Elasticsearch Source.

Authentication Settings

Manual: Enter your credentials directly in the resulting **Username** and **Password** fields.

Secret: Exposes a **Credentials secret** drop-down, in which you can select a stored secret that references the credentials described above. A **Create** link is available to store a new, reusable secret.

Manual API Key: Exposes an **API key** field to directly enter your Elasticsearch API key.

Secret API Key: Exposes an **API key (text secret)** drop-down, in which you can select a stored text secret that references your Elasticsearch API key. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Specify whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Clear Persistent Queue** button. A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process.

Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior.

This setting is required and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. Cribl Edge will append `/<worker-id>/<output-id>` to this field's value.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the `Block` option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Clear Persistent Queue: Click this button if you want to flush out files that are currently queued for delivery to this Destination. A confirmation modal will appear. (Appears only after **Output ID** has been defined.)

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.

- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (s): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Extra parameters: Name-value pairs to pass as additional parameters. If you are using Elastic ingest pipelines, specify an extra parameter whose name is `pipeline` and whose value is the name of your pipeline, similar to these examples.

Include document_id: Toggle this setting to `No` to omit the `document_id` field when sending events to an Elastic TSDS (time series data stream).

Failed request logging mode: Determines which data is logged when a request fails. Use the drop-down to select one of these options:

- `None` (default).
- `Payload`.

- `Payload + Headers`. Use the **Safe Headers** field below to specify the headers to log. If you leave that field empty, all headers are redacted, even with this setting.

Safe headers: List the headers you want to log, in plain text.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Field Normalization

This Destination normalizes the following fields:

- `_time` becomes `@timestamp` at millisecond resolultion.
- `host.name` is set to `host`.

See also our Elasticsearch Source documentation's **Field Normalization** section.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__id`
- `__type`
- `__index`
- `__host`

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.

11.5. Google Cloud

11.5.1. Google Chronicle

Cribl Edge supports sending data to Google Chronicle, a cloud service for retaining, analyzing, and searching enterprise security and network telemetry data.

To define a Google Chronicle Destination, you need to obtain an API key from Google. If you want Cribl Edge or an external KMS to manage the API key, configure a key pair that references the API key.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Chronicle

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Google Cloud** > **Chronicle** Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Google Cloud** > **Chronicle**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Chronicle output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Default log type: Select an application log type to send to Chronicle. (Google Chronicle expects all batches for a given Destination to have the same log type.) Can be overwritten by the `__logType` event field, and interacts with the optional **Custom log types** controls below. See the Google documentation for a current list of supported log types.

Optional Settings

Custom log types: In Cribl Edge 4.0 and later, you can use the **Add type** button here to define each desired type. Each will be a table row with **Log type** and **Description** fields. If you set the **Default log type** drop-down above to the value `Custom`, Cribl Edge will automatically select the first custom log type in this table as the default log type. Use the grab handles at left to reorder table rows.

Namespace: Optionally, configure an environment namespace to identify logs' originating data domain. You can use this as a tag when indexing and/or enriching data. The `__namespace` event field, if present, will overwrite this.

Customer ID: A unique identifier (UUID) corresponding to a particular Chronicle instance. To use this optional field, request the ID from your Chronicle representative.

Send events as: `Unstructured` is the only currently supported format. Cribl plans to add UDM (Unified Data Model) support in a future release.

Log text field: Specify the event field that contains the log text to send. If you do not specify a log text field, Cribl Edge sends a JSON representation of the whole event.

Region: From the drop-down, choose the Google Chronicle regional endpoint to send events to.

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

The Google Chronicle API key is required to complete this part of the Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: In the resulting **API key** field, enter your Google Chronicle API key.
- **Secret**: This option exposes a **Secret** drop-down, in which you can select a stored secret that references your Google Chronicle API key. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This section is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of `KB`, `MB`, etc. Defaults to 1 MB.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of `KB`, `MB`, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

System fields: Specify any fields you want Cribl Edge to automatically add to events using this output. Wildcards are supported. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, `429 (Too Many Requests)` and `503 (Service Unavailable)` are the only response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `30000` (30 seconds).

- **Backoff multiplier**: The base for the exponential backoff algorithm; defaults to `1`. A value of `2` means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Defaults to `30000` (30 seconds); minimum is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Enter an amount of time, in seconds, to wait for a request to complete before aborting it.

Request concurrency: Enter the maximum number of ongoing requests to allow before blocking.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1024` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Enter the maximum number of events to include in the request body. Defaults to `0` (unlimited).

Flush period (sec): Enter the maximum time to allow between requests. Be aware that small values could cause the payload size to be smaller than the configured **Max body size**.

Extra HTTP Headers: Click **Add Header** to insert extra headers as **Name/Value** pairs. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.5.2. Google Cloud Logging

Cribl Edge supports sending log data to Google Cloud Logging, a real-time log storage and management service with search, analysis and alerting.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Google Cloud Logging

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Google Cloud** > **Logging**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Google Cloud** > **Logging**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Google Cloud Logging definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Log location type: Select one of the following.

- `Project`: Displays a **Project ID expression** field. Here, enter a JavaScript expression to compute the value of the project ID to associate log entries with.
- `Organization`: Displays a **Organization ID expression** field. Enter a JS expression to compute the value of the organization ID to associate log entries with.
- `Billing Account`: Displays a **Billing account ID expression** field. Enter a JS expression to compute the value of the billing account ID to associate log entries with.
- `Folder`: Displays a **Folder expression** field. Enter a JS expression to compute the value of the folder ID to associate log entries with.

Log name expression: Ener a JavaScript expression to compute the log name's value.

Field Mappings

Payload format: Select one of the following.

- `Text`: This default option displays a **Payload text expression** field. Here, enter JavaScript expression to compute the payload's value. Must evaluate to a JavaScript string value; if validation fails, this will default to a JSON string representation of the event.
- `JSON`: Displays a **Payload object expression** field. Enter a JS expression to compute the payload's value. Must evaluate to a JavaScript object value; if validation fails, this will default to the entire event.

This Destination compresses all payloads, using gzip.

Log labels: Click **Add label** to apply one or more labels to log entries. On each row of the resulting table, define each label with:

- **Label**: Arbitrary name for the label.
- **Value**: JavaScript expression to compute the label's value.

Resource type expression: JavaScript expression to compute the field value for the managed resource type. Must evaluate to one of the monitored resource types listed here. If not specified, falls back to `global`.

Resource labels: Click **Add label** to apply one or more labels to the managed resource. Each must correspond to a valid label for the resource type you've specified in the **Resource type expression**, or else Google Cloud Logging will drop it. On each row of the resulting table, define each label with:

- **Label**: Arbitrary name for the label.
- **Value**: JavaScript expression to compute the label's value.

Severity expression: JavaScript expression to compute the severity field's value. Must evaluate to one of the `LogSeverity` values listed here. Case-insensitive. If not specified, falls back to `DEFAULT`.

Insert ID expression: JavaScript expression to compute the insert ID field's value.

Optional Settings

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication method** drop-down to select one of these options:

- **Auto**: This option uses the `GOOGLE_APPLICATION_CREDENTIALS` environment variable, and requires no configuration here.
- **Manual**: This default option displays a **Service account credentials** field for you to enter the contents of your service account credentials file (a set of JSON keys), as downloaded from Google Cloud. To insert the file itself, click the upload button at this field's upper right. As an alternative, you can use environment variables, as outlined here.
- **Secret**: This option exposes a drop-down in which you can select a stored secret that references the service account credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish before retrying. Defaults to `10000` (10 sec.).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30` sec.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. Defaults to `0` (unlimited).

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` sec.

Throttle request rate: Enter a maximum number of requests to send per second, as necessary to comply with Google Cloud's API limits. This field defaults to empty (no throttling), and accepts throttling rates as high as `2000` requests per second.

Environment: Optionally, specify a single Git branch on which to enable this configuration. If this field is empty, the config will be enabled everywhere.

Google Cloud Roles and Permissions

Your Google Cloud service account must have at least the following role:

- `Logs Writer`

11.5.3. Google Cloud Pub/Sub

Cribl Edge supports sending data to Google Cloud Pub/Sub, a managed real-time messaging service for sending and receiving messages between applications.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Pub/Sub

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Google Cloud** > **Pub/Sub**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Google Cloud** > **Pub/Sub**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

When working with Google Cloud Pub/Sub, you need to know:

- The project where your credentials originated.
- The project that contains any topic you want to work with.

If the desired topic resides in the same project where your credentials originated, describe it using either its short name **or** its fully-qualified name.

If the desired topic resides in a different project than the one where your credentials originated, you **must** describe it using its fully-qualified name.

Here's an example of a fully-qualified topic name: `projects/my-project-id/topics/my-topic-id`.

The short name for the same topic would be: `my-topic-id`.

Output ID: Enter a unique name to identify this Pub/Sub output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Topic ID: ID of the Pub/Sub topic to send events to. This static initial configuration of the topic is required; but you can optionally also override it dynamically on a per-event basis.

Optional Settings

Create topic: Toggle to `Yes` if you want Cribl Edge to create the topic on Pub/Sub if it does not exist.

Ordered delivery: Toggle to `Yes` if you want Cribl Edge to send events in the order that they arrived in the queue. (For this to work correctly, the process receiving events must have ordering enabled.)

Region: Region to publish messages to. Select `default` to allow Google to auto-select the nearest region. (If you've enabled **Ordered delivery**, the selected region must be allowed by message storage policy.)

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication method** drop-down to select one of these options:

- **Auto**: This option uses the environment variables `PUBSUB_PROJECT` and `PUBSUB_CREDENTIALS`, and requires no configuration here.
- **Manual**: This default option displays a **Service account credentials** field for you to enter the contents of your service account credentials file (a set of JSON keys), as downloaded from Google Cloud.
  To insert the file itself, click the upload button at this field's upper right. As an alternative, you can use environment variables, as outlined here.
- **Secret**: This option exposes a drop-down in which you can select a stored secret that references the service account credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk

> space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.
>
> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Batch size: The maximum number of items the Google API should batch before it sends them to the topic. Defaults to `10` items.

Batch timeout (ms): The maximum interval (in milliseconds) that the Google API should wait to send a batch (if the configured **Batch size** limit has not been reached).. Defaults to `100` ms.

Max queue size: Maximum number of queued batches before blocking. Defaults to `100`.

Max batch size (KB): Maximum size for each sent batch. Defaults to `256` KB.

Max concurrent requests: The maximum number of in-progress API requests before Cribl Edge applies backpressure. Defaults to `10`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Google Cloud Roles and Permissions

Your Google Cloud service account should have at least the following roles on topics:

- `roles/pubsub.publisher`
- `roles/pubsub.viewer` or `roles/viewer`

To enable Cribl Edge's **Create topic** option, your service account should have one of the following (or higher) roles:

- `roles/pubsub.editor`
- `roles/editor`

Either `editor` role confers multiple permissions, including those from the lower `viewer`, `subscriber`, and `publisher` roles. For additional details, see the Google Cloud Access Control topic.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

Let's Change the Topic

The Pub/Sub Destination supports alternate topics specified at the event level in the `__topicOut` field. So (for example) if a Pub/Sub Destination is configured to send to main topic `topic1`, and Cribl Edge receives an event with `__topicOut: topic2`, then Cribl Edge will override the main topic and send this event to `topic2`.

However, a topic specified in the event's `__topicOut` field must already exist on Pub/Sub. If it does not, Cribl Edge cannot dynamically create the topic, and will drop the event. On the Destination's **Status** tab, the **Dropped** metric tracks the number of events dropped because a specified alternate topic did not exist.

(This example uses the short topic name, and that works as long as the topic resides in the project where your GCP credentials originated. If the topic resides in a different project, you must use the fully-qualified topic name, for example `__topicOut: projects/<my-project-id>/topics/topic2`.)

11.5.4. Google Cloud Storage

Google Cloud Storage is a non-streaming Destination type.

Type: Non-Streaming | TLS Support: Yes | PQ Support: No

Configuring Cloud Storage Permissions

For Cribl Edge to send data to Google Cloud Storage buckets, ideally set the following access permissions on the Cloud Storage side:

- Fine-grained access control must be enabled on the buckets.
- The Google service account or user should have the Storage Admin or Owner role.

For narrower access than the above roles, the Google account should have the following access to buckets:

- `roles/storage.objectCreator` role.
- `roles/storage.objectViewer` role.
- If the `Verify if bucket exists` advanced option is enabled, one bucket-level `storage.buckets.list` permission.

For details, see the Cloud Storage Overview of Access Control and Understanding Roles documentation.

Configuring Cribl Edge to Output to Cloud Storage Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles or the **Destinations** left nav, select **Google Cloud** > **Storage**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles, select **Google Cloud** > **Storage**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Cloud Storage definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Bucket name: Name of the destination bucket. This value can be a constant, or a JavaScript expression that can be evaluated only at init time. E.g., referencing a Global Variable: `myBucket-${C.vars.myVar}`.

Region: Region where the bucket is located.

Staging location: Filesystem location in which to locally buffer files before compressing and moving to final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Key prefix: Root directory to prepend to path before uploading. Enter a constant, or a JS expression enclosed in single quotes, double quotes, or backticks.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (supported only on Linux, not Windows) exposes a Parquet Settings left tab, where you **must** configure certain options in order to export data in Parquet format.

Optional Settings

Partitioning expression: JavaScript expression that defines how files are partitioned and organized. Default is date-based. If blank, Cribl Edge will fall back to the event's `__partition` field value (if present); or otherwise to the root directory of the **Output Location** and **Staging Location**.

Compress: Data compression format used before moving to final destination. Defaults to `gzip` (recommended). This setting is not available when **Data format** is set to `Parquet`.

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

To prevent files from being overwritten, Cribl appends a random sequence of 6 characters to the end of their names. File name prefix and suffix expressions do not bypass this behavior.

For example, if you set the **File name prefix expression** to `CriblExec` and set the **File name suffix expression** to `.csv`, the file name might display as `CriblExec-adPRWM.csv` with `adPRWM`

appended.

Backpressure behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication method** drop-down to select one of these options:

- **Auto**: This option authenticates with the attached Google Cloud Platform (GCP) Service Account, relying on GCP IAM roles to access the appropriate GCP resources. This option is available only when Cribl Edge is on-prem, and the Edge Nodes are running in Google Compute Engine VMs on GCP.
- **Manual**: With this default option, authentication is via HMAC (Hash-based Message Authentication Code). To create a key and secret, see Google Cloud's Managing HMAC Keys for Service Accounts documentation. This option exposes these two fields:
  - **Access key**: Enter the HMAC access key.
  - **Secret key**: Enter the HMAC secret.

The values for **Access key** and **Secret key** can be a constant, or a JavaScript expression (such as `${C.env.MY_VAR}`) enclosed in quotes or backticks, which allows configuration with environment variables.

- **Secret**: This option exposes a **Secret key pair** drop-down, in which you can select a stored secret that references the secret key pair described above. A **Create** link is available to store a new, reusable secret.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports `c*` wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported, and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: Set the target memory size for page segments. Generally, set lower values to improve reading speed, or set higher values to improve compression. Value must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files of this size will be closed and moved to final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to final output location. Defaults to `100`.

Cribl Edge will close files when **either** of the `Max file size (MB)` or the `Max file open time (sec)` conditions are met.

Add Output ID: Whether to append output's ID to staging location. Defaults to `Yes`.

Remove staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Endpoint: The Google Cloud Storage service endpoint. Typically, there is no reason to change the default https://storage.googleapis.com endpoint.

Object ACL: Select an Access Control List to assign to uploaded objects. Defaults to `private`.

Storage class: Select a storage class for uploaded objects.

Signature version: Signature version to use for signing requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Troubleshooting

Nonspecific messages from Google Cloud of the form `Error: failed to close file` can indicate problems with the permissions listed above.

11.6. Kafka

11.6.1. Kafka

Cribl Edge supports sending data to a Kafka topic.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Kafka uses a binary protocol over TCP. It does not support HTTP proxies, so Cribl Edge must send events directly to receivers. You might need to adjust your firewall rules to allow this traffic.

Configuring Cribl Edge to Output to Kafka

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Kafka**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Kafka**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Kafka definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Brokers: List of Kafka brokers to connect to. (E.g., `localhost:9092`.)

Topic: The topic on which to publish events. Can be overwritten using event's `__topicOut` field.

Optional Settings

Acknowledgments: Select the number of required acknowledgments. Defaults to `Leader`.

Record data format: Format to use to serialize events before writing to Kafka. Defaults to `JSON`. When set to `Protobuf`, the **Protobuf Format Settings** section (left tab) becomes available.

Compression: Codec to compress the data before sending to Kafka. Select `None`, `Gzip` `Snappy`, or `LZ4`.

Cribl strongly recommends enabling compression. Doing so improves Cribl Edge's performance, enabling faster data transfer using less bandwidth.

Backpressure behavior: Select whether to block, drop, or queue incoming events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Select `None`, `Gzip`, `Snappy`, or `LZ4`. Defaults to `Gzip`.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down.**Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Calculating the Time PQ Will Take to Engage

PQ will not engage until Cribl Edge has exhausted all attempts to send events to the Kafka receiver. This can take several minutes if requests continue to fail or time out.

To calculate the longest possible time this can take, multiply the values of **Advanced Settings** > **Request timeout** and **Max retries**. For the default values (`60` seconds and `5`, respectively), this would be `60` seconds times `5` retries = 300 seconds, or 5 minutes.

TLS Settings (Client Side)

Enabled Defaults to `No`. When toggled to `Yes`:

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Leave this field blank. See Connecting to Kafka below.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Authentication

This section documents the SASL (Simple Authentication and Security Layer) authentication settings to use when connecting to brokers. Using TLS is highly recommended.

Enabled: Defaults to `No`. When toggled to `Yes`:

SASL mechanism: Use this drop-down to select the SASL authentication mechanism to use. The mechanism you select determines the controls displayed below.

PLAIN, SCRAM-256, or SCRAM-512

With any of these SASL mechanisms, select one of the following buttons:

Manual: Displays **Username** and **Password** fields to enter your Kafka credentials directly.

Secret: This option exposes a **Credentials secret** drop-down in which you can select a stored text secret that references your Kafka credentials. A **Create** link is available to store a new, reusable secret.

GSSAPI/Kerberos

Selecting Kerberos as the authentication mechanism displays the following options:

Keytab location: Enter the location of the keytab file for the authentication principal.

Principal: Enter the authentication principal, e.g.: `kafka_user@example.com`.

Broker service class: Enter the Kerberos service class for Kafka brokers, e.g.: `kafka`.

You will also need to set up your environment and configure the Cribl Stream host for use with Kerberos. See Kafka Authentication with Kerberos for further detail.

Schema Registry

This section governs Kafka Schema Registry Authentication for Avro-encoded data with a schema stored in the Confluent Schema Registry.

Enabled: defaults to `No`. When toggled to `Yes`, displays the following controls:

Schema registry URL: URL for access to the Confluent Schema Registry. (E.g., `http://<hostname>:8081`.)

Default key schema ID: Used when `__keySchemaIdOut` is not present to transform key values. Leave blank if key transformation is not required by default.

Default value schema ID: Used when `__valueSchemaIdOut` not present to transform `_raw`. Leave blank if value transformation is not required by default.

TLS enabled: defaults to `No`. When toggled to `Yes`, displays the following TLS settings for the Schema Registry (in the same format as the TLS Settings (Client Side) above):

- **Validate server certs**: Require client to reject any connection that is not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Server name (SNI)**: Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.
- **Minimum TLS version**: Optionally, select the minimum TLS version to use when connecting.
- **Maximum TLS version**: Optionally, select the maximum TLS version to use when connecting.
- **Certificate name**: The name of the predefined certificate.
- **CA certificate path**: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.
- **Private key path (mutual auth)**: Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.
- **Certificate path (mutual auth)**: Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.
- **Passphrase**: Passphrase to use to decrypt private key.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (KB, uncompressed): Maximum size (KB) of each record batch before compression. Setting should be < `message.max.bytes settings` in Kafka brokers. Defaults to `768`.

Max events per batch: Maximum number of events in a batch before forcing a flush. Defaults to `1000`.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Connection timeout (ms): Maximum time to wait for a successful connection. Defaults to `10000` ms, i.e., 10 seconds. Valid range is `1000` to `3600000` ms, i.e., 1 second to 1 hour.

Request timeout (ms): Maximum time to wait for a successful request. Defaults to `60000` ms, i.e., 1 minute.

Max retries: Maximum number of times to retry a failed request before the message fails. Defaults to `5`; enter `0` to not retry at all.

Authentication timeout (ms): Maximum time to wait for Kafka to respond to an authentication request. Defaults to `1000` (1 second).

Reauthentication threshold (ms): If the broker requires periodic reauthentication, this setting defines how long before the reauthentication timeout Cribl Edge initiates the reauthentication. Defaults to `10000` (10 seconds).

A small value for this setting, combined with high network latency, might prevent the Destination from reauthenticating before the Kafka broker closes the connection.

A large value might cause the Destination to send reauthentication messages too soon, wasting bandwidth.

The Kafka setting `connections.max.reauth.ms` controls the reuthentication threshold on the Kafka side.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Protobuf Format Settings

Definition set: From the drop-down, select `OpenTelemetry`. This makes the **Object type** setting available.

Object type: From the drop-down, select `Logs`, `Metrics`, or `Traces`.

Working with Protobufs

This Destination supports supports Binary Protobuf payload encoding. The Protobufs it sends can encode traces, metrics, or logs, the three types of telemetry data defined in the OpenTelemetry Project's Data Sources documentation:

- A **trace** tracks the progression of a single request.
  - Each trace is a tree of **spans**.
  - A span object represents the work being done by the individual services, or components, involved in a request as that request flows through a system.
- A **metric** provides aggreggated statistical information.
  - A metric contains individual measurements called **data points**.
- A **log**, in OpenTelemetry terms, is "any data that is not part of a distributed trace or a metric".

When configuring Pipelines (including pre-processing and post-processing Pipelines), you need to ensure that events sent to this Destination conform to the relevant Protobuf specification:

- For traces, opentelemetry/proto/trace/v1/trace.proto.
- For metrics, opentelemetry/proto/metrics/v1/metrics.proto.
- For logs, opentelemetry/proto/logs/v1/logs.proto.

This Destination will drop non-conforming events. If the Destination encounters a parsing error when trying to convert an event to OTLP, it discards the event and Cribl Edge logs the error.

Connecting to Kafka

Leave the TLS Settings > Server name (SNI) field blank

In Cribl Edge's Kafka-based Sources and Destinations (including this one), the Kafka library that Cribl Edge uses manages SNI (Server Name Indication) without any input from Cribl Edge. Therefore, you should leave the **TLS Settings** > **Server name (SNI)** field blank.

Setting this field in the Cribl Edge UI can cause traffic to be routed to the wrong brokers, because it interferes with the Kafka library's operation.

Connecting to a Kafka cluster entails working with hostnames for **brokers** and **bootstrap servers**.

Brokers are servers that comprise the storage layer in a Kafka cluster. Bootstrap servers handle the initial connection to the Kafka cluster, and then return the list of brokers. A broker list can run into the hundreds. Every Kafka cluster has a `bootstrap.servers` property, defined as either a single `hostname:port` K-V pair, or a list of them. If Cribl Edge tries to connect via one bootstrap server and that fails, Cribl Edge then tries another one on the list.

In the **General Settings** > **Brokers** list, you can enter either the hostnames of brokers that your Kafka server has been configured to use, or, the hostnames of one or more bootstrap servers. If Kafka returns a list of brokers that's longer than the list you entered, Cribl Edge keeps the full list internally. Cribl Edge neither saves the list nor makes it available in the UI. The connection process simply starts at the beginning whenever the Source or Destination is started.

Here's an overview of the connection process:

1. From the **General Settings** > **Brokers** list – where each broker is listed as a hostname and port – Cribl Edge takes a hostname and resolves it to an IP address.
2. Cribl Edge makes a connection to that IP address. Notwithstanding the fact that Cribl Edge resolved **one** particular hostname to that IP address, there may be **many** services running at that IP address – each with its own distinct hostname.
3. Cribl Edge establishes TLS security for the connection.

Although SNI is managed by the Kafka library rather than in the Cribl Edge UI, you might want to know how it fits into the connection process. The purpose of the SNI is to specify one hostname – i.e., service – among many that might be running on a given IP address within a Kafka cluster. Excluding the other services is one way that TLS makes the connection more secure.

Working with Event Timestamps

By default, when an incoming event reaches this Destination, the underlying Kafka JS library adds a `timestamp` field set to the current time at that moment. The library gets the current time from the `Date.now()` JavaScript method. Events that this Destination sends to downstream Kafka receivers include this `timestamp` field.

Some production situations require the `timestamp` value to be the time an incoming event was originally created by an upstream sender, not the current time of its arrival at this Destination. For example, suppose the events were originally created over a wide time range, and arrive at the Destination hours or days later. In this case, the original creation time might be important to retain in events Cribl Edge sends to downstream Kafka receivers.

Here is how to satisfy such a requirement:

1. In a Pipeline that routes events to this Destination, use an Eval Function to add a field named `__kafkaTime` and write the incoming event's original creation time into that field. The value of `__kafkaTime` **must** be in UNIX epoch time (either seconds or milliseconds since the UNIX epoch – both will work). For context, see this explanation of the related fields `_time` and `__origTime`.
2. This Destination will recognize the `__kafkaTime` field and write its value into the `timestamp` field. (This is similar to the way you can use the `__topicOut` field to overwrite a topic setting, as described above.)

If the `__kafkaTime` field is **not** present, the Destination will apply the default behavior (using `Date.now()`) described previously.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__topicOut`
- `__key`
- `__headers`
- `__keySchemaIdOut`
- `__valueSchemaIdOut`

11.6.2. Confluent Cloud

Cribl Edge supports sending data to Kafka topics on the Confluent Cloud managed Kafka platform.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Confluent Cloud uses a binary protocol over TCP. It does not support HTTP proxies, so Cribl Edge must send events directly to receivers. You might need to adjust your firewall rules to allow this traffic.

Sending Kafka Topic Data to Confluent Cloud

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Confluent Cloud**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Confluent Cloud**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition.

Brokers: List of Confluent Cloud brokers to connect to. (E.g., `myAccount.confluent.cloud:9092`.)

Topic: The topic on which to publish events. Can be overwritten using event's `__topicOut` field.

Optional Settings

Acknowledgments: Select the number of required acknowledgments. Defaults to `Leader`.

Record data format: Format to use to serialize events before writing to Kafka. Defaults to `JSON`. When set to `Protobuf`, the **Protobuf Format Settings** section (left tab) becomes available.

Compression: Codec to use to compress the data before sending to Kafka. Select `None`, `Gzip`, `Snappy`, or `LZ4`. Defaults to `Gzip`.

Backpressure behavior: Select whether to block, drop, or queue incoming events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip`, `Snappy`, and `LZ4` are also available.

Cribl strongly recommends enabling compression. Doing so improves Cribl Edge's performance, enabling faster data transfer using less bandwidth.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Calculating the Time PQ Will Take to Engage

PQ will not engage until Cribl Edge has exhausted all attempts to send events to the Kafka receiver. This can take several minutes if requests continue to fail or time out.

To calculate the longest possible time this can take, multiply the values of **Advanced Settings** > **Request timeout** and **Max retries**. For the default values (`60` seconds and `5`, respectively), this would be `60` seconds times `5` retries = 300 seconds, or 5 minutes.

TLS Settings (Client Side)

Enabled When toggled to `Yes` (the default):

Autofill?: This setting is experimental.

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Server name (SNI): Leave this field blank. See Connecting to Kafka below.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Authentication

This section documents the SASL (Simple Authentication and Security Layer) authentication settings to use when connecting to brokers. Using TLS is highly recommended.

Enabled: Defaults to `No`. When toggled to `Yes`:

- **SASL mechanism**: Select the SASL (Simple Authentication and Security Layer) authentication mechanism to use. Defaults to `PLAIN`. `SCRAM-SHA-256`, `SCRAM-SHA-512`, and `GSSAPI/Kerberos` are also available. The mechanism you select determines the controls displayed below.

PLAIN, SCRAM-256, or SCRAM-512

With any of these SASL mechanisms, select one of the following buttons:

Manual: Displays **Username** and **Password** fields to enter your Kafka credentials directly.

Secret: This option exposes a **Credentials secret** drop-down in which you can select a stored text secret that references your Kafka credentials. A **Create** link is available to store a new, reusable secret.

GSSAPI/Kerberos

Selecting Kerberos as the authentication mechanism displays the following options:

Keytab location: Enter the location of the keytab file for the authentication principal.

Principal: Enter the authentication principal, e.g.: `kafka_user@example.com`.

Broker service class: Enter the Kerberos service class for Kafka brokers, e.g.: `kafka`.

You will also need to set up your environment and configure the Cribl Stream host for use with Kerberos. See Kafka Authentication with Kerberos for further detail.

Schema Registry

This section governs Kafka Schema Registry Authentication for Avro-encoded data with a schema stored in the Confluent Schema Registry.

Enabled: Defaults to `No`. When toggled to `Yes`, displays the following controls:

Schema registry URL: URL for access to the Confluent Schema Registry. (E.g., `http://localhost:8081`.)

Default key schema ID: Used when `__keySchemaIdOut` is not present to transform key values. Leave blank if key transformation is not required by default.

Default value schema ID: Used when `__valueSchemaIdOut` not present to transform `_raw`. Leave blank if value transformation is not required by default.

TLS enabled: defaults to `No`. When toggled to `Yes`, displays the following TLS settings for the Schema Registry (in the same format as the TLS Settings (Client Side) above):

- **Validate server certs**: Require client to reject any connection that is not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `No`.
- **Server name (SNI)**: Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

  With a dedicated Confluent Cloud cluster hosted in Microsoft Azure, be sure to specify the **Server name (SNI)**. If this is omitted, Confluent Cloud might reset the connection to Cribl Edge.

- **Minimum TLS version**: Optionally, select the minimum TLS version to use when connecting.
- **Maximum TLS version**: Optionally, select the maximum TLS version to use when connecting.
- **Certificate name**: The name of the predefined certificate.
- **CA certificate path**: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.
- **Private key path (mutual auth)**: Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.
- **Certificate path (mutual auth)**: Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.
- **Passphrase**: Passphrase to use to decrypt private key.

Processing Settings

Post-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data before it is sent through this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (KB, uncompressed): Maximum size (KB) of each record batch before compression. Setting should be < `message.max.bytes settings` in Kafka brokers. Defaults to `768`.

Max events per batch: Maximum number of events in a batch before forcing a flush. Defaults to `1000`.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Connection timeout (ms): Maximum time to wait for a successful connection. Defaults to `10000` ms, i.e., 10 seconds. Valid range is `1000` to `3600000` ms, i.e., 1 second to 1 hour.

Request timeout (ms): Maximum time to wait for a successful request. Defaults to `60000` ms, i.e., 1 minute.

Max retries: Maximum number of times to retry a failed request before the message fails. Defaults to `5`; enter `0` to not retry at all.

Authentication timeout (ms): Maximum time to wait for Kafka to respond to an authentication request. Defaults to `1000` (1 second).

Reauthentication threshold (ms): If the broker requires periodic reauthentication, this setting defines how long before the reauthentication timeout Cribl Edge initiates the reauthentication. Defaults to `10000` (10 seconds).

A small value for this setting, combined with high network latency, might prevent the Destination from reauthenticating before the Kafka broker closes the connection.

A large value might cause the Destination to send reauthentication messages too soon, wasting bandwidth.

The Kafka setting `connections.max.reauth.ms` controls the reuthentication threshold on the Kafka side.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Protobuf Format Settings

Definition set: From the drop-down, select `OpenTelemetry`. This makes the **Object type** setting available.

Object type: From the drop-down, select `Logs`, `Metrics`, or `Traces`.

Working with Protobufs

This Destination supports supports Binary Protobuf payload encoding. The Protobufs it sends can encode traces, metrics, or logs, the three types of telemetry data defined in the OpenTelemetry Project's Data Sources documentation:

- A **trace** tracks the progression of a single request.
  - Each trace is a tree of **spans**.
  - A span object represents the work being done by the individual services, or components, involved in a request as that request flows through a system.
- A **metric** provides aggreggated statistical information.
  - A metric contains individual measurements called **data points**.
- A **log**, in OpenTelemetry terms, is "any data that is not part of a distributed trace or a metric".

When configuring Pipelines (including pre-processing and post-processing Pipelines), you need to ensure that events sent to this Destination conform to the relevant Protobuf specification:

- For traces, opentelemetry/proto/trace/v1/trace.proto.
- For metrics, opentelemetry/proto/metrics/v1/metrics.proto.
- For logs, opentelemetry/proto/logs/v1/logs.proto.

This Destination will drop non-conforming events. If the Destination encounters a parsing error when trying to convert an event to OTLP, it discards the event and Cribl Edge logs the error.

Connecting to Kafka

Leave the TLS Settings > Server name (SNI) field blank

In Cribl Edge's Kafka-based Sources and Destinations (including this one), the Kafka library that Cribl Edge uses manages SNI (Server Name Indication) without any input from Cribl Edge. Therefore, you should leave the **TLS Settings** > **Server name (SNI)** field blank.

Setting this field in the Cribl Edge UI can cause traffic to be routed to the wrong brokers, because it interferes with the Kafka library's operation. This is especially important for Confluent Cloud Dedicated clusters, which rely on SNI – as managed by the Kafka library – for routing.

Connecting to a Kafka cluster entails working with hostnames for **brokers** and **bootstrap servers**.

Brokers are servers that comprise the storage layer in a Kafka cluster. Bootstrap servers handle the initial connection to the Kafka cluster, and then return the list of brokers. A broker list can run into the hundreds. Every Kafka cluster has a `bootstrap.servers` property, defined as either a single `hostname:port` K-V pair, or a list of them. If Cribl Edge tries to connect via one bootstrap server and that fails, Cribl Edge then tries another one on the list.

In the **General Settings** > **Brokers** list, you can enter either the hostnames of brokers that your Kafka server has been configured to use, or, the hostnames of one or more bootstrap servers. If Kafka returns a list of brokers that's longer than the list you entered, Cribl Edge keeps the full list internally. Cribl Edge neither saves the list nor makes it available in the UI. The connection process simply starts at the beginning whenever the Source or Destination is started.

Here's an overview of the connection process:

1. From the **General Settings** > **Brokers** list – where each broker is listed as a hostname and port – Cribl Edge takes a hostname and resolves it to an IP address.
2. Cribl Edge makes a connection to that IP address. Notwithstanding the fact that Cribl Edge resolved **one** particular hostname to that IP address, there may be **many** services running at that IP address – each with its own distinct hostname.
3. Cribl Edge establishes TLS security for the connection.

Although SNI is managed by the Kafka library rather than in the Cribl Edge UI, you might want to know how it fits into the connection process. The purpose of the SNI is to specify one hostname – i.e., service – among many that might be running on a given IP address within a Kafka cluster. Excluding the other services is one way that TLS makes the connection more secure.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__topicOut`
- `__key`
- `__headers`
- `__keySchemaIdOut`
- `__valueSchemaIdOut`

11.6.3. Amazon MSK

Cribl Edge supports sending data to an Amazon Managed Streaming for Apache Kafka (MSK) topic.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Kafka uses a binary protocol over TCP. It does not support HTTP proxies, so Cribl Edge must send events directly to receivers. You might need to adjust your firewall rules to allow this traffic.

Configuring Cribl Edge to Output to Kafka

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Amazon MSK**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Amazon MSK**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Amazon MSK Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Brokers: List of Kafka brokers to connect to. (E.g., `kafkaBrokerHost:9092`.)

Topic: The topic on which to publish events. Can be overwritten using the event's `__topicOut` field.

Region: From the drop-down, select the name of the AWS Region where your Amazon MSK cluster is located.

Optional Settings

Acknowledgments: Select the number of required acknowledgments. Defaults to `Leader`.

Record data format: Format to use to serialize events before writing to Kafka. Defaults to `JSON`. When set to `Protobuf`, the **Protobuf Format Settings** section (left tab) becomes available.

Compression: Codec to compress the data before sending to Kafka. Select `None`, `Gzip`, `Snappy`, or `LZ4`. Defaults to `Gzip`.

Cribl strongly recommends enabling compression. Doing so improves Cribl Edge's performance, enabling faster data transfer using less bandwidth.

Backpressure behavior: Select whether to block, drop, or queue incoming events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block**

option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Calculating the Time PQ Will Take to Engage

PQ will not engage until Cribl Edge has exhausted all attempts to send events to the Kafka receiver. This can take several minutes if requests continue to fail or time out.

To calculate the longest possible time this can take, multiply the values of **Advanced Settings** > **Request timeout** and **Max retries**. For the default values (`60` seconds and `5`, respectively), this would be `60` seconds times `5` retries = 300 seconds, or 5 minutes.

TLS Settings (Client Side)

For Amazon MSK Sources and Destinations:

- IAM is the only type of authentication that Cribl Edge supports.
- Because IAM auth requires TLS, TLS is automatically enabled.

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Leave this field blank. See Connecting to Kafka below.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Authentication

Use the **Authentication method** drop-down to select an AWS authentication method.

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. The **Secret** option exposes this additional field:

- **Secret key pair**: Use the drop-down to select an API key/secret key pair that you've configured in Cribl Edge's secrets manager. A **Create** link is available to store a new, reusable secret.

Assume Role

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION`

environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

Enable for MSK: Toggle on to use Assume Role credentials to access MSK.

AssumeRole ARN: Enter the Amazon Resource Name (ARN) of the role to assume.

External ID: Enter the External ID to use when assuming role. This is required only when assuming a role that requires this ID in order to delegate third-party access. For details, see AWS' documentation.

Duration (seconds): Duration of the Assumed Role's session, in seconds. Minimum is 900 (15 minutes). Maximum is 43200 (12 hours). Defaults to 3600 (1 hour).

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (KB, uncompressed): Maximum size (KB) of each record batch before compression. Setting should be < `message.max.bytes settings` in Kafka brokers. Defaults to `768`.

Max events per batch: Maximum number of events in a batch before forcing a flush. Defaults to `1000`.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Connection timeout (ms): Maximum time to wait for a connection to complete successfully. Defaults to `10000` ms, i.e., 10 seconds. Valid range is `1000` to `3600000` ms, i.e., 1 second to 1 hour.

Request timeout (ms): Maximum time to wait for Kafka to respond to a request. Defaults to `60000` ms, i.e., 1 minute.

Max retries: Maximum number of times to retry a failed request before the message fails. Defaults to `5`; enter `0` to not retry at all.

Authentication timeout (ms): Maximum time to wait for Kafka to respond to an authentication request. Defaults to `1000` (1 second).

Reauthentication threshold (ms): If the broker requires periodic reauthentication, this setting defines how long before the reauthentication timeout Cribl Edge initiates the reauthentication. Defaults to `10000` (10 seconds).

A small value for this setting, combined with high network latency, might prevent the Destination from reauthenticating before the Kafka broker closes the connection.

A large value might cause the Destination to send reauthentication messages too soon, wasting bandwidth.

The Kafka setting `connections.max.reauth.ms` controls the reuthentication threshold on the Kafka side.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Protobuf Format Settings

Definition set: From the drop-down, select `OpenTelemetry`. This makes the **Object type** setting available.

Object type: From the drop-down, select `Logs`, `Metrics`, or `Traces`.

Working with Protobufs

This Destination supports supports Binary Protobuf payload encoding. The Protobufs it sends can encode traces, metrics, or logs, the three types of telemetry data defined in the OpenTelemetry Project's Data Sources documentation:

- A **trace** tracks the progression of a single request.
  - Each trace is a tree of **spans**.
  - A span object represents the work being done by the individual services, or components, involved in a request as that request flows through a system.
- A **metric** provides aggreggated statistical information.
  - A metric contains individual measurements called **data points**.

- A **log**, in OpenTelemetry terms, is "any data that is not part of a distributed trace or a metric".

When configuring Pipelines (including pre-processing and post-processing Pipelines), you need to ensure that events sent to this Destination conform to the relevant Protobuf specification:

- For traces, opentelemetry/proto/trace/v1/trace.proto.
- For metrics, opentelemetry/proto/metrics/v1/metrics.proto.
- For logs, opentelemetry/proto/logs/v1/logs.proto.

This Destination will drop non-conforming events. If the Destination encounters a parsing error when trying to convert an event to OTLP, it discards the event and Cribl Edge logs the error.

Connecting to Kafka

Leave the TLS Settings > Server name (SNI) field blank

In Cribl Edge's Kafka-based Sources and Destinations (including this one), the Kafka library that Cribl Edge uses manages SNI (Server Name Indication) without any input from Cribl Edge. Therefore, you should leave the **TLS Settings** > **Server name (SNI)** field blank.

Setting this field in the Cribl Edge UI can cause traffic to be routed to the wrong brokers, because it interferes with the Kafka library's operation.

Connecting to a Kafka cluster entails working with hostnames for **brokers** and **bootstrap servers**.

Brokers are servers that comprise the storage layer in a Kafka cluster. Bootstrap servers handle the initial connection to the Kafka cluster, and then return the list of brokers. A broker list can run into the hundreds. Every Kafka cluster has a `bootstrap.servers` property, defined as either a single `hostname:port` K-V pair, or a list of them. If Cribl Edge tries to connect via one bootstrap server and that fails, Cribl Edge then tries another one on the list.

In the **General Settings** > **Brokers** list, you can enter either the hostnames of brokers that your Kafka server has been configured to use, or, the hostnames of one or more bootstrap servers. If Kafka returns a list of brokers that's longer than the list you entered, Cribl Edge keeps the full list internally. Cribl Edge neither saves the list nor makes it available in the UI. The connection process simply starts at the beginning whenever the Source or Destination is started.

Here's an overview of the connection process:

1. From the **General Settings** > **Brokers** list – where each broker is listed as a hostname and port – Cribl Edge takes a hostname and resolves it to an IP address.

2. Cribl Edge makes a connection to that IP address. Notwithstanding the fact that Cribl Edge resolved **one** particular hostname to that IP address, there may be **many** services running at that IP address – each with its own distinct hostname.
3. Cribl Edge establishes TLS security for the connection.

Although SNI is managed by the Kafka library rather than in the Cribl Edge UI, you might want to know how it fits into the connection process. The purpose of the SNI is to specify one hostname – i.e., service – among many that might be running on a given IP address within a Kafka cluster. Excluding the other services is one way that TLS makes the connection more secure.

Working with Event Timestamps

By default, when an incoming event reaches this Destination, the underlying Kafka JS library adds a `timestamp` field set to the current time at that moment. The library gets the current time from the `Date.now()` JavaScript method. Events that this Destination sends to downstream Kafka receivers include this `timestamp` field.

Some production situations require the `timestamp` value to be the time an incoming event was originally created by an upstream sender, not the current time of its arrival at this Destination. For example, suppose the events were originally created over a wide time range, and arrive at the Destination hours or days later. In this case, the original creation time might be important to retain in events Cribl Edge sends to downstream Kafka receivers.

Here is how to satisfy such a requirement:

1. In a Pipeline that routes events to this Destination, use an Eval Function to add a field named `__kafkaTime` and write the incoming event's original creation time into that field. The value of `__kafkaTime` **must** be in UNIX epoch time (either seconds or milliseconds since the UNIX epoch – both will work). For context, see this explanation of the related fields `_time` and `__origTime`.
2. This Destination will recognize the `__kafkaTime` field and write its value into the `timestamp` field. (This is similar to the way you can use the `__topicOut` field to overwrite a topic setting, as described above.)

If the `__kafkaTime` field is **not** present, the Destination will apply the default behavior (using `Date.now()`) described previously.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__topicOut`
- `__key`
- `__headers`
- `__keySchemaIdOut`
- `__valueSchemaIdOut`

11.7. Metrics

11.7.1. Graphite

Cribl Edge supports sending data to a Graphite backend Destination.

Type: Streaming | TLS Support: No | PQ Support: Yes

Configuring Cribl Edge to Output to a Graphite Backend

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Metrics** > **Graphite**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Metrics** > **Graphite**.
Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Graphite definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Destination protocol: Protocol to use when communicating with the Destination. Defaults to `UDP`.

Host: The hostname of the Destination.

Port: Destination port. Defaults to `8125`.

Optional Settings

Throttling: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Rate (in bytes per second) at which at which to throttle while writing to an output. Also takes numerical values in multiples of bytes (KB, MB, GB, etc.). Default value of `0` indicates no throttling.

Backpressure behavior: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`, and only when **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Timeout Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`.

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000`.

Write timeout: Amount of time (milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (bytes): Used when Protocol is UDP. Specifies the maximum size of packets sent to the Destination. (Also known as the MTU – maximum transmission unit – for the network path to the destination system.) Defaults to `512`.

Flush period (sec): Used when Protocol is TCP. Specifies how often buffers should be flushed, sending records to the Destination. Defaults to `1`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.7.2. StatsD

Cribl Edge supports sending data to a StatsD Destination.

Type: Streaming | TLS Support: No | PQ Support: Yes

Configuring Cribl Edge to Output via StatsD

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Metrics** > **StatsD**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Metrics** > **StatsD**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this StatsD definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Destination protocol: Protocol to use when communicating with the Destination. Defaults to `UDP`.

Host: The hostname of the Destination.

Port: Destination port. Defaults to `8125`.

Optional Settings

Throttling: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Rate (in bytes per second) at which at which to throttle while writing to an output. Also takes numerical values in multiples of bytes (KB, MB, GB, etc.). Default value of `0` indicates no throttling.

Backpressure behavior: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`, and only when **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue fallback behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** drops the newest events from being sent out of Cribl Edge, and throws away incoming data, while leaving the contents of the PQ unchanged.

Timeout Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`.

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000`.

Write timeout: Amount of time (milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (bytes): Used when Protocol is UDP. Specifies the maximum size of packets sent to the Destination. (Also known as the MTU – maximum transmission unit – for the network path to the Destination system.) Defaults to `512`.

Flush period (sec): Used when Protocol is TCP. Specifies how often buffers should be flushed, sending records to the Destination. Defaults to `1`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.7.3. StatsD Extended

Cribl Edge's StatsD Extended Destination supports sending out data in expanded StatsD format.

The output is an expanded StatsD metric protocol that supports dimensions, along with a sample rate for counter metrics. As with StatsD, downstream components listen for application metrics over UDP or TCP, can aggregate and summarize those metrics, and can relay them to virtually any graphing or monitoring backend.

For details about the syntax expected by one common downstream service, see Splunk's Expanded StatsD Metric Protocol documentation.

Type: Streaming | TLS Support: No | PQ Support: Yes

Configuring Cribl Edge to Output via StatsD Extended

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Metrics** > **StatsD Extended**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Metrics** > **StatsD Extended**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this StatsD Extended definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Destination protocol: Protocol to use when communicating with the Destination. Defaults to `UDP`.

Host: The hostname of the Destination.

Port: Destination port. Defaults to `8125`.

Optional Settings

Throttling: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Rate (in bytes per second) at which at which to throttle while writing to an output. Also takes numerical values in multiples of bytes (KB, MB, GB, etc.). Default value of `0` indicates no throttling.

Backpressure behavior: Displayed only when **General Settings** > **Destination protocol** is set to `TCP`. Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`, and only when **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue fallback behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** drops the newest events from being sent out of Cribl Edge, and throws away incoming data, while leaving the contents of the PQ unchanged.

Timeout Settings

This section is displayed only when **General Settings** > **Destination protocol** is set to `TCP`.

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000`.

Write timeout: Amount of time (milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Max record size (bytes): Used when Protocol is UDP. Specifies the maximum size of packets sent to the Destination. (Also known as the MTU – maximum transmission unit – for the network path to the Destination system.) Defaults to `512`.

Flush period (sec): Used when Protocol is TCP. Specifies how often buffers should be flushed, sending records to the Destination. Defaults to `1`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.8. New Relic Ingest

11.8.1. New Relic Events

Cribl Edge supports sending events to New Relic via the New Relic Event API. Use this Destination to export ad hoc events that New Relic ingestion treats as custom events.

To export structured log and/or metric events, use Cribl Edge's New Relic Logs & Metrics Destination.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output Events to New Relic

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **New Relic Ingest** > **Events**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **New Relic Ingest** > **Events**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination.

Region: Select which New Relic region endpoint to use.

Account ID: Enter your New Relic account ID. (You can access this ID from New Relic's **account** drop-down, by selecting **Manage your plan**.)

Event type: Default `eventType` to apply when not specified in an event. You can use arbitrary values, as long as they do not conflict with New Relic reserved words. (Where an `eventType` is specified in an event, it will override this value.)

Authentication Settings

If an incoming event contains an internal field named `__newRelic_apiKey`, the New Relic Events Destination will use that field's value as the API key when sending the event to New Relic.

For events that do not contain a `__newRelic_apiKey` field, the Destination will use whatever API key you have configured in this section.

Authentication method: Select one of the following buttons.

- **Manual**: This default option exposes an **API key** field. Directly enter your New Relic Ingest License API key, as you created or accessed it from New Relic's **account** drop-down. (For details, see the New Relic API Keys documentation.)
- **Secret**: This option exposes an **API key (text secret)** drop-down, in which you can select a stored secret that references a New Relic Ingest License API key. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`. For the `Persistent Queue` option, see the section just below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed when the **Backpressure behavior** is set to `Persistent Queue`.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out via this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Defaults to `Yes`, meaning compress the payload body before sending.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1024` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. Defaults to `0`, allowing unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` second.

Extra HTTP headers: Click **Add Header** to insert extra headers as **Name**/**Value** pairs. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Verifying the New Relic Events Destination

Once you've configured event sources, create one or more Routes to send data to New Relic.

In New Relic, you can create visualizations incorporating the Cribl Edge-supplied data, then add them to new or existing dashboards as widgets.

Alternatively, in the New Relic backend, you can select **Query you data** (top nav) > **Events** (left tab), and then select the event type you exported from Cribl Edge.

To view more events, change the time frame at the upper right. To see raw events, click **Raw data** on the right.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.8.2. New Relic Logs & Metrics

Cribl Edge supports sending events to the New Relic Log API and the New Relic Metric API.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

In Cribl Edge v.3.1.2 and later, this Destination now authenticates using New Relic's Ingest License API key. (New Relic will retire the Insights Insert API keys, which this Destination previously used for authentication.)

Also in v.3.1.2 and later, Cribl Edge provides a separate New Relic Events Destination that you can use to send ad hoc (loosely structured) events to New Relic via the New Relic Event API.

Within New Relic's platform, you can monitor Cribl Edge's performance and data flow by installing New Relic's Cribl dashboard.

Configuring Cribl Edge to Output to New Relic

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **New Relic Ingest** > **Logs & Metrics**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **New Relic Ingest** > **Logs & Metrics**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this New Relic definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Region: Select which New Relic region endpoint to use.

Authentication Settings

If an incoming event contains an internal field named `__newRelic_apiKey`, the New Relic Logs & Metrics Destination uses that field's value as the API key when sending the event to New Relic.

> For events that do not contain an `__newRelic_apiKey` field, the Destination uses whatever API key you have configured in the **Authentication method** settings.

Authentication method: Select one of the following buttons.

- **Manual**: This default option exposes an **API key** field. Directly enter your New Relic Ingest License API key, as you created or accessed it from New Relic's **account** drop-down. (For details, see the New Relic API Keys documentation.)
- **Secret**: This option exposes an **API key (text secret)** drop-down, in which you can select a stored secret that references a New Relic Ingest License API key. A **Create** link is available to store a new, reusable secret.

Optional Settings

Log type: Name of the `logType` to send with events. E.g., `observability` or `access_log`.

> This sets a default. Where a `sourcetype` is specified in an event, it will override this value.

Log message field: Name of the field to send as the log `message` value. If not specified, the event will be serialized and sent as JSON.

Fields: Additional fields to (optionally) add, as **Name**-**Value** pairs. Click **Add Field** to add more.

- **Name**: Enter the field name.
- **Value**:JavaScript expression to compute field's value, enclosed in single quotes, double quotes, or backticks. (Can evaluate to a constant.)

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`. For the `Persistent Queue` option, see the section just below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

> This section is displayed when the **Backpressure behavior** is set to `Persistent Queue`.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other

options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).

- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1024` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` second.

Extra HTTP headers: Click **Add Header** to insert extra headers as **Name/Value** pairs. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Verifying the New Relic Destination

Once you've configured log and/or metrics sources, create one or more Routes to send data to New Relic.

In New Relic, you can create visualizations incorporating the Cribl Edge-supplied data, then add them to new or existing dashboards as widgets.

Logs and metrics land in two different places in New Relic.

Log Queries

To access and query log data:

- Navigate to the New Relic home screen's **Logs** header option, and click the **(+)** button at right.
- Then to build your queries, use the **Find logs where** input field, and add desired columns to the table view below the graph,.

Metrics Queries

To access and query metrics data:

- From the New Relic home screen, *Click **Browse Data** > **Metrics** > **Can Search for metricNames**.
- Then, customize time range and dimensions to build the desired logic for your queries.
- Alternatively, you can use NRQL to build your own query searches.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.9. Prometheus

11.9.1. Prometheus

Cribl Edge can send metric events to targets and third-party platforms that support Prometheus' remote write specification (overview here).

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output to Prometheus

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Prometheus**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Prometheus**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Prometheus output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Remote Write URL: The endpoint to send events to, e.g.: `http://localhost:9200/write`

Optional Settings

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.
- **Auth token (text secret)**: This option exposes a **Token (text secret)** drop-down, in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block**

option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_host` (Cribl Edge Node that processed the event) and `cribl_wp` (Cribl Edge Worker Process that processed the event). Supports wildcards. Other options include:

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_pipe` – Cribl Edge Pipeline that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are **not** authorized by a trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Metric renaming expression: A JavaScript expression that can be used to rename metrics.
The default expression – `name.replace(/\\./g, \'_\')` – replaces all `.` characters in a metric's name with the Prometheus-supported `_` character. Use the `name` global variable to access the metric's name. You can access event fields' values via `__e.<fieldName>`.

Send metadata: Whether to generate and send metrics' metadata (`type` and `metricFamilyName`) along with the metrics. The default `Yes` value displays this additional field:

- **Metadata flush period (sec)**: How frequently metrics metadata is sent out. Value must at least equal the base **Flush period (sec)**. (In other words, metadata cannot be flushed on a shorter interval.) Defaults to `60` seconds.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header

names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

If an event contains the internal field `__criblMetrics`, Cribl Edge will send it to the HTTP endpoint as a metric event. Otherwise, Cribl Edge will drop the event.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Unlike other HTTP-based Destinations, Prometheus does not display an **Advanced Settings** > **Compress** option. The Prometheus `remote_write` spec assumes that payloads are snappy-compressed by default.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between Prometheus cluster nodes.

11.9.2. Grafana Cloud

Cribl Edge can send data to two of the services available in Grafana Cloud: Loki for logs and Prometheus for metrics. The Grafana Cloud Destination shapes events appropriately for Loki and Prometheus, and routes events to the correct endpoint for each service.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Preparing Prometheus and Loki to Receive Data from Cribl Edge

To define a Grafana Cloud Destination, you need a Grafana Cloud account.

While logged in to your Grafana account, navigate to the Grafana Cloud Portal, which should be located at `https://grafana.com/orgs/<your-organization-name>`, and complete the following steps.

Obtain an API key, setting its Role to `MetricsPublisher`. If you want Cribl Edge or an external KMS to manage the API key, configure a key pair that references the API key.

In the Prometheus tile, click **Send Metrics** to open the Prometheus configuration page. Write down:

- Your **Remote Write Endpoint** URL, for example: `https://prometheus-blocks-prod-us-central1.grafana.net/api/prom/push`.
- Your Prometheus **Username**.

In the Loki tile, click **Send Logs** to open the Loki configuration page. Write down:

- Your **Grafana Data Source settings** URL, for example: `https://logs-prod-us-central1.grafana.net`.
- Your Loki **User** ID.

Decide what type of authentication to use and prepare accordingly:

- If you choose Basic authentication, the username (**Username** in Prometheus, **User** in Loki) and password (simply your Grafana API key) will remain separate.
- If you choose token-based authentication, construct your tokens by concatenating username, colon (`:`), and password, for example `12345:cOQvDj6sJGFS3Bk2MguBW==`. Because the Prometheus and Loki usernames differ, you need to construct a separate token for each service.

Configuring Cribl Edge to Output to Grafana Cloud

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Grafana Cloud**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Grafana Cloud**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Grafana Cloud output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Loki URL: The endpoint to send log events to, e.g.: `https://logs-prod-us-central1.grafana.net`. This is the **Grafana Data Source settings** URL you wrote down earlier.

Prometheus URL: The endpoint to send metric events to, e.g.: `https://prometheus-blocks-prod-us-central1.grafana.net/api/prom/push`. This is the **Remote Write Endpoint** URL you wrote down earlier.

Optional Settings

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`. `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Authentication

The **Authentication** tab provides separate **Loki** and **Prometheus** sections, enabling you to configure these inputs separately. The two sections provide identical options.

Select one of the following options for authentication:

- **Auth token**: Enter the bearer token that must be included in the authorization header. Use the token that you constructed earlier. In Grafana Cloud, the bearer token is generally built by concatenating the

username and the API key, separated by a colon. E.g.: `<your-username>:<your-api-key>`.

- **Auth token (text secret)**: This option exposes a drop-down in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.
- **Basic**: This default option displays fields for you to enter HTTP Basic authentication credentials. **Username** is the Loki **User** or Prometheus **Username** that you wrote down earlier. **Password** is your API key in the Grafana Cloud domain.
- **Basic (credentials secret)**: This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.

Processing Settings

Metric events can have dimensions, and log events have labels. Dimensions, labels, and their values are determined by several different settings in Cribl Edge. This section explains how that works, along with other kinds of settings.

Loki uses labels to define separate streams of logging data. This is a key concept. Cribl recommends that you familiarize yourself with the information and documentation Grafana provides about labels in Loki.

One canonical example is processing logs from servers in three environments: production, staging, and testing. You could create a label named `env` whose possible values are `prod`, `staging`, and `test`.

One basic principle is that if you set too many labels, you can end up with too many streams.

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output—both metric events, as dimensions; and, log events, as labels. Supports wildcards.

By default, includes `cribl_host` (Cribl Edge Node that processed the event) and `cribl_wp` (Cribl Edge Worker Process that processed the event). On the Loki side, this creates different streams, which prevents Loki from rejecting some events as being out of order when different Nodes or Worker Processes are emitting at different rates.

Other options include:

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.

- `cribl_pipe` – Cribl Edge Pipeline that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Loki and Prometheus might complain about entries being delivered out of order when **Request concurrency** is set > `1` and any of **Flush period (sec)**, **Max body size (KB)**, or **Max events per request** are set to low values.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Metric renaming expression: A JavaScript expression that can be used to rename metrics.
The default expression – `name.replace(/\\./g, \'_\')` – replaces all `.` characters in a metric's name with the Prometheus-supported `_` character. Use the `name` global variable to access the metric's name. You can access event fields' values via `__e.<fieldName>`.

Message format: Whether to send events as `Protobuf` (the default) or `JSON`.

Compress: When the **Message format** is `JSON`, this setting controls the data compression format used before sending the data to Grafana Cloud. Defaults to `Yes` for GZIP-compression. (Applies only to Loki's JSON payloads. This toggle is hidden when the **Message format** is `Protobuf`, because both Prometheus' and Loki's Protobuf implementations are Snappy-compressed by default.)

Logs message field: The event field to send as log output, for example: `_raw`. All other event fields are discarded. If left blank, Cribl Edge sends a JSON representation of the whole event.

Logs labels: Name/value pairs where the value can be a static or dynamic expression that has access to all log event fields.

Failed request logging mode: Determines which data is logged when a request fails. Use the drop-down to select one of these options:

- `None` (default).
- `Payload`.
- `Payload + Headers`. Use the **Safe Headers** field below to specify the headers to log. If you leave that field empty, all headers are redacted, even with this setting.

Safe headers: List the headers you want to log, in plain text.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

If an event contains the internal field `__criblMetrics`, Cribl Edge will send it Prometheus as a metric event. If `__criblMetrics` is absent, Cribl Edge will treat the event as a log and send it to Loki.

The internal field `__labels` specifies labels to add to log events. If a label is set in both the `__labels` field and in **Logs labels** and/or **System fields**, Cribl Edge sends the value from `__labels` to Loki. Setting the

`__labels` field in a Pipeline gives you a quick way to experiment with the logs being sent.

If there are no labels set (this would happen when **System fields**, **Logs labels**, and `__labels` are all empty), Cribl Edge adds a default `source` label, which prevents Loki from rejecting events. The `source` label the concatenation of `cribl`, underscore (`_`), source type, colon (`:`), source-name, where source name and type are values in the `__inputId` event field, for example: `cribl_metrics:in_prometheus_rw`. If `__inputId` is missing, `source` is set to `cribl`.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- The **Advanced Settings** > **Compress** toggle determines whether to compress the payload body before sending to Loki only. The toggle setting does not apply to Prometheus payloads, which are always compressed using Snappy.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between Grafana Cloud nodes.

11.9.3. Loki

Cribl Edge can send log events to Grafana's Loki log aggregation system.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Loki

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Loki**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Loki**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Loki output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Loki URL: The endpoint to send events to, e.g.: `https://logs-prod-us-central1.grafana.net`.

To proxy outbound HTTP/S requests, see System Proxy Configuration.

Optional Settings

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.
- **Auth token (text secret)**: This option exposes a drop-down in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.
- **Basic**: This default option displays fields for you to enter HTTP Basic authentication credentials. **Username** is the Loki **User**. **Password** is your API key in the Grafana Cloud domain.
- **Basic (credentials secret)**: This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`. `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block**

option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Loki uses labels to define separate streams of logging data. This is a key concept. Cribl recommends that you familiarize yourself with the information and documentation Grafana provides about labels in Loki.

One canonical example is processing logs from servers in three environments: production, staging, and testing. You could create a label named `env` whose possible values are `prod`, `staging`, and `test`.

One basic principle is that if you set too many labels, you can end up with too many streams.

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to log events as labels. Supports wildcards.

By default, includes `cribl_host` (Cribl Edge Node that processed the event) and `cribl_wp` (Cribl Edge Worker Process that processed the event). On the Loki side, this creates different streams, which prevents Loki from rejecting some events as being out of order when different Nodes or Worker Processes are emitting at different rates.

Other options include:

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.

- `cribl_pipe` – Cribl Edge Pipeline that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Validate server certs: Reject certificates that are not authorized by a trusted CA (e.g., the system's CA). Defaults to `Yes`.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `1`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `15`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Message format: Whether to send events as `Protobuf` (the default) or `JSON`.

Compress: When the **Message format** is `JSON`, this setting controls the data compression format used before sending the data to Loki. Defaults to `Yes` for GZIP-compression. (Applies only to Loki's JSON payloads.

This toggle is hidden when the **Message format** is `Protobuf`, because both Prometheus' and Loki's Protobuf implementations are Snappy-compressed by default.)

Logs message field: The event field to send as log output, for example: `_raw`. All other event fields are discarded. If left blank, Cribl Edge sends a JSON or Protobuf representation of the whole event.

Logs labels: Name/value pairs where the value can be a static or dynamic expression that has access to all log event fields.

Failed request logging mode: Determines which data is logged when a request fails. Use the drop-down to select one of these options:

- `None` (default).
- `Payload`.
- `Payload + Headers`. Use the **Safe Headers** field below to specify the headers to log. If you leave that field empty, all headers are redacted, even with this setting.

Safe headers: List the headers you want to log, in plain text.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.10. Splunk

11.10.1. Splunk HEC

The **Splunk HEC** Destination can stream data to a Splunk HEC (HTTP Event Collector) receiver through the event endpoint. The data arrives to Splunk cooked and parsed, so it enters at the Splunk data pipeline's indexing segment.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Events sent to the Splunk HEC Destination will show higher outbound data volume than the same events sent to the Splunk Single Instance or Splunk Load Balanced Destinations, which use the S2S binary protocol.

Configuring Cribl Edge to Output to Splunk HEC Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Splunk** > **HEC**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Splunk** > **HEC**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Splunk HEC definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Load balancing: When enabled (default), lets you specify multiple Splunk HEC endpoints and load weights. With the default `No` setting, if you notice that Cribl Edge is not sending data to all possible IP addresses, enable **Advanced Settings** > **Round-robin DNS**.

Splunk HEC endpoint: URL of a Splunk HEC endpoint to send events to (e.g., `http://myhost.example.com:8088/services/collector/event`). This setting appears only

when **Load balancing** is toggled to `No`.

Splunk HEC Endpoints

Use the **Splunk HEC Endpoints** table to specify a known set of receivers on which to load-balance data. It appears only when **Load balancing** is toggled to `Yes`.

To specify more receivers on new rows, click **Add Endpoint**. Each row provides the following fields:

HEC Endpoint: Specify the URL to a Splunk HEC endpoint to send events to – e.g., `http://localhost:8088/services/collector/event`.

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

When you first enable load balancing, or if you edit the load weight once your data is load–balanced, give the logic time to settle. The changes might take a few seconds to register.

For details on configuring all these options, see About Load Balancing.

For Splunk Cloud endpoints, change the **Splunk HEC endpoint**'s default `http:` prefix to: `https:`.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays an **HEC Auth token** field for you to enter your Splunk HEC API access token.
- **Secret**: This option exposes an **HEC Auth token (text secret)** drop-down, in which you can select a stored secret that references the API access token described above. A **Create** link is available to store a new, reusable secret.

This Destination does not support Mutual TLS (mTLS).

Indexer Acknowledgement

Do not set **Enable Indexer Acknowledgement** on the Splunk token. With this setting enabled, Splunk receivers expect the Channel GUID to be passed in, and requests will fail with errors of this form:

```
channel: output:splunk_cloud_hec
cid: w2
level: error
message: Request failed
reason: Received status code=400, method=POST, url=https://http-inputs-bazookatron
response: {"text":"Data channel is missing","code":10}
time: 2023-02-14T15:26:03.413Z
```

Optional Settings

Exclude current host IPs: This toggle appears when **Load balancing** is set to `Yes`. It determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).

- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Output multi-metrics: Toggle to `Yes` to output multiple-measurement metric data points. (Supported in Splunk 8.0 and above, this format enables sending multiple metrics in a single event, improving the efficiency of your Splunk capacity.)

Validate server certs: Reject certificates that are not authorized by a trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned. (This setting is available only when **General Settings** > **Load balancing** is set to `No`.)

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`. Each request can potentially hit a different HEC receiver.

Max body size (KB): Maximum size, in KB, of the request body. Defaults to `4096`. Lowering the size can potentially result in more parallel requests and also cause outbound requests to be made sooner.

If you're sending data to Splunk Cloud, Cribl recommends a maximum value of `1 MB` for **Max body size (KB)**. This upper limit is defined by the maximum content length for the default HTTP Event Collector (HEC) provided by Splunk Cloud.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1`.

- Retries happen on this flush interval.
- Any HTTP response code in the `2xx` range is considered success.
- Any response code in the `5xx` range is considered a retryable error, which will not trigger Persistent Queue (PQ) usage.
- Any other response code will trigger PQ (if PQ is configured as the Backpressure behavior).

Extra HTTP headers: Click **Add Header** to add **Name**/**Value** pairs to pass as additional HTTP headers. Values will be sent encrypted.

The next two fields take effect only in the Cribl App for Splunk. (Cribl Edge ignores their values.)

Next processing queue: Specify the next Splunk processing queue to send the events to, after HEC processing. Defaults to `indexQueue`.

Default _TCP_ROUTING: Specify the value of the `_TCP_ROUTING` field for events that do not have `_ctrl._TCP_ROUTING` set. Defaults to `nowhere`. This is useful only when you expect the HEC receiver to route this data on to another destination.

The next two fields appear only when the **General Settings** > **Load balancing** option is set to `Yes`.

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that if multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Edge load balancing, IP settings take priority over those from hostnames.)

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new connection will be reattempted. This is to prevent sticking to a particular Destination when there is a constant flow of events.
- If the server does not support keepalives – or if the server closes a pooled connection while idle – a new connection will be established for the next request.
- When resolving the Destination's hostname with load balancing disabled, Cribl Stream will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between Splunk HEC servers.
- See Splunk's documentation on editing `fields.conf` to ensure the visibility of index-time fields sent to Splunk by Cribl Edge.

Troubleshooting Resources

Cribl University offers an Advanced Troubleshooting > Destination Integrations: Splunk Cloud short course. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

11.10.2. Splunk Single Instance

This Splunk Destination can stream data to a **free** Splunk Cloud instance. From the perspective of the receiving Splunk Cloud instance, the data arrives cooked and parsed.

For a **Standard** Splunk Cloud instance whose `../default/outputs.conf` file contains multiple indexer entries, you must instead use Cribl Edge's Splunk Load Balanced Destination.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output to Splunk Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles or the **Destinations** left nav, select **Splunk** > **Single Instance**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles, select **Splunk** > **Single Instance**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Splunk Single Instance definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Address: Hostname of the Splunk receiver.

Port: The port number on the host.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** drops the newest events from being sent out of Cribl Edge, and throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

TLS Settings (Client Side)

Enabled defaults to `No`. When toggled to `Yes`:

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Single .pem File

If you have a **single** .pem file containing `cacert`, `key`, and `cert` sections, enter it in all of these fields above: **CA certificate path**, **Private key path (mutual auth)**, and **Certificate path (mutual auth)**.

Timeout Settings

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000`.

Write timeout: Amount of time (in milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Output multiple metrics: Toggle to `Yes` to output multiple-measurement metric data points. (Supported in Splunk 8.0 and above, this format enables sending multiple metrics in a single event, improving the efficiency of your Splunk capacity.)

Minimize in-flight data loss: If set to `Yes` (the default), Cribl Edge will check whether the indexer is shutting down, and if so, will stop sending data. This helps minimize data loss during shutdown. (Note that Splunk logs will indicate that the Cribl app has set `UseAck` to `true`, even though Cribl does not enable full `UseAck` behavior.) If toggled to `No`, exposes the following alternative option:

Max failed health checks: Displayed (and set to `1` by default) only if **Minimize in-flight data loss** is disabled. This option sends periodic requests to Splunk once per minute, to verify that the Splunk endpoint is still alive and can receive data. Its value governs how many failed requests Cribl Edge will allow before closing this connection.

A low threshold value improves connections' resilience, but by proliferating connections, this can complicate troubleshooting. Set to `0` to disable health checks entirely – here, if the connection to

> Splunk is forcibly closed, you risk some data loss.

Max S2S version: The highest version of the Splunk-to-Splunk protocol to expose during handshake. Defaults to `v3; v4` is also available.

Throttling: Throttle rate, in bytes per second. Defaults to `0`, meaning no throttling. Multiple-byte units such as KB, MB, GB etc. are also allowed, e.g., `42 MB`. When throttling is engaged, your **Backpressure behavior** selection determines whether Cribl Edge will handle excess data by blocking it, dropping it, or queueing it to disk.

Nested field serialization: Specifies how to serialize nested fields into index-time fields. Defaults to `None`.

Authentication method: Use the buttons to select one of these options:

- **Manual**: In the resulting **Auth token** field, enter the shared secret token to use when establishing a connection to a Splunk indexer.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the auth token described above. A **Create** link is available to store a new, reusable secret.

Log failed requests to disk: Toggling to `Yes` makes the payloads of any (future) failed requests available for inspection. See Inspecting Payloads to Troubleshoot Closed Connections below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Inspecting Payloads to Troubleshoot Closed Connections

When a downstream receiver closes connections from this Destination (or just stops responding), inspecting the payloads of the resulting failed requests can help you find the cause. For example:

- Suppose you send an event whose size is larger than the downstream receiver can handle.
- Suppose you send an event that has a `number` field, but the value exceeds the highest number that the downstream receiver can handle.

When **Log failed requests to disk** is enabled, you can inspect the payloads of failed requests. Here is how:

1. In the Destination UI, navigate to the **Logs** tab.
2. Find a log entry with a `connection error` message.
3. Expand the log entry.

4. If the message includes the phrase `See payload file for more info`, note the path in the `file` field on the next line.

Now you have the path to the directory where Cribl Edge is storing payloads from failed requests. At the command line, navigate to that directory and inspect any payloads that you think might be relevant.

Notes about Forwarding to Splunk

- Data sent to Splunk is not compressed.
- The only `ack` from indexers that Cribl Edge listens for and acts upon is the shutdown signal described in Minimize in-flight data loss above.
- If events have a Cribl Edge internal field called `__criblMetrics`, they'll be forwarded to Splunk as metric events.
- If events do **not** have a `_raw` field, they'll be serialized to JSON prior to sending to Splunk.
- See Splunk's documentation on editing `fields.conf` to ensure the visibility of index-time fields sent to Splunk by Cribl Edge.

Troubleshooting Resources

Cribl University offers an Advanced Troubleshooting > Destination Integrations: Splunk Cloud short course. To follow the direct course link, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

11.10.3. Splunk Load Balanced

The **Splunk Load Balanced** Destination can load-balance the data it streams to multiple Splunk receivers. Downstream Splunk instances receive the data cooked and parsed.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

For additional details about sending to Splunk Cloud, see Splunk Cloud and BYOL Integrations.

Configuring Cribl Edge to Load-Balance to Multiple Splunk Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Splunk** > **Load Balanced**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Splunk** > **Load Balanced**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Splunk LB Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Toggling **Indexer discovery** to `Yes` enables automatic discovery of indexers in an indexer clustering environment. This hides both **Exclude current host IPs** and the **Destinations** section, and displays the following fields:

Site: Clustering site from which indexers need to be discovered. In the case of a single site cluster, `default` is the default entry.

Cluster Manager URI: Full URI of Splunk cluster manager, in the format: `scheme://host:port`. (Worker Nodes/Edge Nodes normally access the cluster manager on **port 8089** to get the list of currently online indexers.)

Refresh period: Time interval (in seconds) between two consecutive fetches of indexer list from cluster manager. Defaults to `300` seconds, i.e., 5 minutes.

Authentication method: Use the buttons to select one of these options for authenticating to cluster Manager for indexer discovery:

- **Manual**: In the resulting **Auth token** field, enter the required token.
- **Secret**: This option exposes a **Auth token (text secret)** drop-down, in which you can select a stored secret that references the auth token. A **Create** link is available to store a new, reusable secret.

Each Worker Process performs its own indexer discovery according to the above settings.

Destinations

The **Destinations** section appears only when **Indexer discovery** is set to its `No` default. Here, you specify a known set of Splunk receivers on which to load-balance data.

Click **Add Destination** to specify more receivers on new rows. Each row provides the following fields:

- **Address**: Hostname of the Splunk receiver. Optionally, you can paste in a comma-separated list, in `<host>:<port>` format.
- **Port**: Port number to send data to.
- **TLS**: Whether to inherit TLS configs from group setting, or disable TLS. Defaults to `inherit`.
- **TLS servername**: Servername to use if establishing a TLS connection. If not specified, defaults to connection host (if not an IP). Otherwise, uses the global TLS settings.
- **Load weight**: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the **Destinations** table.

For details on configuring all load balancing settings, see About Load Balancing.

Optional Settings

Toggle **Exclude current host IPs** to `Yes` if you want to exclude all the current host's IP addresses from the list of resolved hostnames.

Backpressure behavior: Select whether to block, drop, or queue events when all receivers in this group are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`. When toggled to `Persistent Queue`, adds the Persistent Queue Settings section (left tab) to the modal.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Enabling Cluster Manager Authentication

To enable token authentication on the Splunk cluster manager, you can find complete instructions in Splunk's Enable or Disable Token Authentication documentation. This option requires Splunk 7.3.0 or higher, and requires the following capabilites: `list_indexer_cluster` and `list_indexerdiscovery`.

For details on creating the token, see Splunk's Create Authentication Tokens topic – especially its section on how to Configure Token Expiry and "Not Before" Settings.

Be sure to give the token an **Expiration** setting well in the future, whether you use **Relative Time** or **Absolute Time**. Otherwise, the token will inherit Splunk's default expiration time of `+30d` (30 days in the future), which will cause indexer discovery to fail.

If you have a failover site configured on Splunk's cluster manager, Cribl respects this configuration, and forwards the data to the failover site in case of site failure.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down.. **Drop new data** drops the newest events being sent out of Cribl Edge, throws away incoming data, and leaves the contents of the PQ unchanged.

TLS Settings (Client Side)

Enabled defaults to `No`. When toggled to `Yes`:

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Single PEM File

If you have a **single** `.pem` file containing `cacert`, `key`, and `cert` sections, enter this file's path in all of these fields above: **CA certificate path**, **Private key path (mutual auth)**, and **Certificate path (mutual auth)**.

Timeout Settings

- **Connection timeout**: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000` ms.
- **Write timeout**: Amount of time (in milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000` ms.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Output multiple metrics: Toggle to `Yes` to output multiple-measurement metric data points. (Supported in Splunk 8.0 and above, this format enables sending multiple metrics in a single event, improving the efficiency of your Splunk capacity.)

Minimize in-flight data loss: If set to `Yes` (the default), Cribl Edge will check whether the indexer is shutting down, and if so, will stop sending data. This helps minimize data loss during shutdown. (Note that Splunk logs will indicate that the Cribl app has set `UseAck` to `true`, even though Cribl does not enable full `UseAck` behavior.) If toggled to `No`, exposes the following alternative option:

Max failed health checks: Displayed (and set to `1` by default) only if **Minimize in-flight data loss** is disabled. This option sends periodic requests to Splunk once per minute, to verify that the Splunk endpoint is still alive and can receive data. Its value governs how many failed requests Cribl Edge will allow before closing this connection.

A low threshold value improves connections' resilience, but by proliferating connections, this can complicate troubleshooting. Set to `0` to disable health checks entirely – here, if the connection to Splunk is forcibly closed, you risk some data loss.

Max S2S version: The highest version of the Splunk-to-Splunk protocol to expose during handshake. Defaults to `v3`; `v4` is also available.

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that If multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Edge load balancing, IP settings take priority over those from hostnames.)

Max connections: Constrains the number of concurrent indexer connections, per Worker Process, to limit memory utilization. If set to a number > `0`, then on every DNS resolution period (or indexer discovery), Cribl Edge will randomly select this subset of discovered IPs to connect to. Cribl Edge will rotate IPs in future resolution periods – monitoring weight and historical data, to ensure fair load balancing of events among IPs.

Nested field serialization: Specifies whether and how to serialize nested fields into index-time fields. Select `None` (the default) or `JSON`.

Authentication method: Use the buttons to select one of these options:

- **Manual**: In the resulting **Auth token** field, enter the shared secret token to use when establishing a connection to a Splunk indexer.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the auth token described above. A **Create** link is available to store a new, reusable secret.

Log failed requests to disk: Toggling to `Yes` makes the payloads of any (future) failed requests available for inspection. See Inspecting Payloads to Troubleshoot Closed Connections below.

Endpoint health fluctuation time allowance (ms): How long (in milliseconds) each receiver endpoint can report blocked, before this Destination as a whole reports unhealthy, blocking senders. (Grace period for

transient fluctuations.) Use `0` to disable the allowance; default is `100` ms; maximum allowed value is `60000` ms (1 minute).

Throttling: Throttle rate, in bytes per second. Multiple byte units such as KB, MB, GB, etc., are also allowed. E.g., `42 MB`. Default value of `0` indicates no throttling. When throttling is engaged, excess data will be dropped only if **Backpressure behavior** is set to **Drop events**. (Data will be blocked for all other **Backpressure behavior** settings.)

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Inspecting Payloads to Troubleshoot Closed Connections

When a downstream receiver closes connections from this Destination (or just stops responding), inspecting the payloads of the resulting failed requests can help you find the cause. For example:

- Suppose you send an event whose size is larger than the downstream receiver can handle.
- Suppose you send an event that has a `number` field, but the value exceeds the highest number that the downstream receiver can handle.

When **Log failed requests to disk** is enabled, you can inspect the payloads of failed requests. Here is how:

1. In the Destination UI, navigate to the **Logs** tab.
2. Find a log entry with a `connection error` message.
3. Expand the log entry.
4. If the message includes the phrase `See payload file for more info`, note the path in the `file` field on the next line.

Now you have the path to the directory where Cribl Edge is storing payloads from failed requests. At the command line, navigate to that directory and inspect any payloads that you think might be relevant.

SSL Configuration for Splunk Cloud – Special Note

To connect to Splunk Cloud, you will need to extract the private and public keys from the Splunk-provided Splunk Cloud Universal Forwarder credentials package. You will also need to reference the CA Certificate located in the same package.

You can reuse many of the settings in this Splunk Cloud package to set up Splunk Cloud Destinations. Use the following steps:

Step 1. Extract the `splunkclouduf.spl` package on the Cribl Edge instance that you will be connecting to Splunk Cloud. You will have a folder that looks something like this:

```
100_my-splunk-cloud_splunkcloud
        /default/
                outputs.conf
                limits.conf
                your-splunk-cloud_server.pem
                your-splunk-cloud_cacert.pem
```

Step 2. (optional) Test connectivity to Splunk Cloud, using the Root CA certificate:

```
echo | openssl s_client -CAfile 100_<your-splunk-cloud>_splunkcloud/default/my-splu
```

To test the connection, you can use any of the URLs listed in the `[tcpout:splunkcloud]` stanza's `outputs.conf` section.

You can simplify Steps 3 and 4 below by dragging and dropping (or uploading) the `.pem` files into Cribl Edge's **New Certificates** modal. See SSL Certificate Configuration.

Step 3. Extract the private key from the Splunk Cloud certificate. At the prompt, you will need the `sslPassword` value from the `outputs.conf`. Using Elliptic Curve keys:

```
openssl ec -in 100_<your-splunk-cloud>_splunkcloud/default/<your-splunk-cloud>_serv
```

If you are using RSA keys, instead use:

```
openssl rsa -in 100_<your-splunk-cloud>_splunkcloud/default/<your-splunk-cloud>_se
```

Step 4. Extract the public key for the Server Certificate:

```
openssl x509 -in 100_<your-splunk-cloud>_splunkcloud/default/<your-splunk-cloud>_se
```

Step 5. In the Splunk Load Balanced Destination's TLS Settings (Client Side) section, enter the following:

- CA Certificate Path: Path to `<your-splunk-cloud>_cacert.pem`.
- Private Key Path (mutual auth): Path to `private.pem` (Step 3 above).

- Certificate Path (mutual auth): Path to `server.pem` (Step 4 above).

In a distributed deployment, enter this Destination configuration on each Worker Group/Fleet that forwards to Splunk Cloud. Then commit and deploy your changes.

Step 6. In a distributed deployment, enable Worker UI access, and verify that the Certificate files have been distributed to individual workers. If they are not present, copy the Certificate files to the Workers, using exactly the same paths you used at the Group level.

Notes About Forwarding to Splunk

- Data sent to Splunk is **not** compressed.
- The only `ack` from indexers that Cribl Edge listens for and acts upon is the shutdown signal described in Minimize in-flight data loss above.
- If events have a Cribl Edge internal field called `__criblMetrics`, they'll be forwarded to Splunk as metric events.
- If events do not have a `_raw` field, they'll be serialized to JSON prior to sending to Splunk.
- You can copy and paste the Splunk Cloud servers from the `[tcpout:splunkcloud]` stanza into the Splunk Load Balanced Destination's **General Settings** > **Destinations** section. E.g., from the example stanza below, you would copy only the bolded contents:

  ~~[tcpout:splunkcloud]~~
  ~~server =~~ **inputs1.your-splunk-cloud.splunkcloud.com:9997, inputs2.your-splunk-cloud.splunkcloud.com:9997, inputs3.your-splunk-cloud.splunkcloud.com:9997, inputs4.your-splunk-cloud.splunkcloud.com:9997, inputs5.your-splunk-cloud.splunkcloud.com:9997, inputs6.your-splunk-cloud.splunkcloud.com:9997, inputs7.your-splunk-cloud.splunkcloud.com:9997, inputs8.your-splunk-cloud.splunkcloud.com:9997, inputs9.your-splunk-cloud.splunkcloud.com:9997, inputs10.your-splunk-cloud.splunkcloud.com:9997, inputs11.your-splunk-cloud.splunkcloud.com:9997, inputs12.your-splunk-cloud.splunkcloud.com:9997, inputs13.your-splunk-cloud.splunkcloud.com:9997, inputs14.your-splunk-cloud.splunkcloud.com:9997, inputs15.your-splunk-cloud.splunkcloud.com:9997**
  ~~compressed = false~~
  From `limits.conf`, copy the `[thruput]` value, and paste it into the Splunk Load Balanced Destination's **Advanced Settings** tab > **Throttling** setting.
- If you enable TLS including cert validation, indexer discovery might trigger errors. This is because by default, Cribl will get the indexers' IPs from their certs, not their fully qualified domain names (FQDNs). As a workaround, use `server.conf` on each indexer, setting `register_forwarder_address = <your.idx.fqdn>`. Cribl will now get that value, and the certs will match.
- See Splunk's documentation on editing `fields.conf` to ensure the visibility of index-time fields sent to Splunk by Cribl Edge.

Troubleshooting Resources

Cribl University's Advanced Troubleshooting short courses include Destination Integrations: Splunk LB and Destination Integrations: Splunk Cloud. To follow these direct course links, first log into your Cribl University account. (To create an account, click the **Sign up** link. You'll need to click through a short **Terms & Conditions** presentation, with chill music, before proceeding to courses – but Cribl's training is always free of charge.) Once logged in, check out other useful Advanced Troubleshooting short courses and Troubleshooting Criblets.

11.11. CrowdStrike Falcon LogScale

The **CrowdStrike Falcon LogScale** Destination can stream data to a LogScale HEC (HTTP Event Collector) in JSON or Raw format.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

(In Cribl Edge 3.5.x, this Destination was labeled **Humio HEC**. Some links from this page might still lead to "Humio"-branded resources that CrowdStrike has not renamed.)

Recommendations

To load-balance to customer-managed LogScale receivers, Cribl recommends placing a load balancer in front of cluster nodes, following LogScale Manual Cluster Deployment recommendations. If you are using LogScale Cloud, it provides its own load balancing.

We recommend sending events with the `sourceType` field set to a LogScale parser. This tells LogScale which parser to use to extract fields (e.g., `"sourceType":"json"`).

If LogScale cannot match the `sourceType` value to a parser, it will use the `kv` parser, and you will get an error that LogScale could not resolve the specified parser. Alternatively, you can assign a parser to the ingest token that you use to authenticate this Destination.

The `fields` element does not support Nested JSON. Any nested elements will be dropped.

Configuring Cribl Edge to Output to CrowdStrike Falcon LogScale Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **CrowdStrike Falcon LogScale**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **CrowdStrike Falcon LogScale**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this CrowdStrike Falcon LogScale definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

LogScale endpoint: URL of a CrowdStrike Falcon LogScale endpoint to send events to (e.g., `https://cloud.us.humio.com:443/api/v1/ingest/hec`).

- JSON-formatted events normally go to `/api/v1/ingest/hec` or `/services/collector`.
- Raw (simple line-delimited) events normally go to `/api/v1/ingest/hec/raw` or `/services/collector/raw`.

Request format: Select how you want events formatted, either `JSON` or `Raw`. Make sure your selection here matches the **LogScale endpoint** you specify above.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays a **LogScale Auth token** field for you to enter your CrowdStrike Falcon LogScale HEC API token.
- **Secret**: Displays a **LogScale token (text secret)** drop-down, in which you can select a stored secret that references the API token described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.

- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or

3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Defaults to `Yes` to reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA).

Round–robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned. (This setting is available only when **General Settings** > **Load balancing** is set to `No`.)

Compress: Defaults to `Yes` to compress the payload body before sending.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`. Each request can potentially hit a different HEC receiver.

Max body size (KB): Maximum size, in KB, of the request body. Defaults to `4096`. Lowering the size can potentially result in more parallel requests and also cause outbound requests to be made sooner.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1`.

- Retries happen on this flush interval.

- Any HTTP response code in the `2xx` range is considered success.
- Any response code in the `5xx` range is considered a retryable error, which will not trigger Persistent Queue (PQ) usage.
- Any other response code will trigger PQ (if PQ is configured as the Backpressure behavior).

Extra HTTP headers: Click **Add Header** to add **Name**/**Value** pairs to pass as additional HTTP headers. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers that you want to declare as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new connection will be reattempted. This is to prevent sticking to a particular Destination when there is a constant flow of events.
- If the server does not support keepalives – or if the server closes a pooled connection while idle – a new connection will be established for the next request.
- Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between CrowdStrike Falcon LogScale servers.

11.12. Datadog

Cribl Edge can send log and metric events to Datadog. (Datadog supports metrics only of type `gauge`, `counter`, and `rate` via its REST API.)

Cribl Edge sends events to the following Datadog endpoints in the US region. Use a DNS lookup to discover and include the corresponding IP addresses in your firewall rules' allowlist.

- Logs: `https://http-intake.logs.datadoghq.com/v1/input`
- Metrics: `https://api.datadoghq.com/api/v1/series`

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Datadog

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Datadog**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Datadog**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays a field for you to enter an **API key** that is available in your Datadog profile.
- **Secret**: This option exposes an **API key (text secret)** drop-down, in which you can select a stored secret that references the API access token described above. A **Create** link is available to store a new, reusable secret.

API key: Enter your Datadog organization's API key.

Optional Settings

Datadog site: Select the Datadog region you are sending to. Defaults to `US`; the other options are `US3`, `US5`, `Europe`, `US1-FED`, and `AP1`.

Send logs as: Specify the content type to use when sending logs. Defaults to `application/json`, where each log message is represented by a JSON object. The alternative `text/plain` option sends one message per line, with newline `\n` delimiters.

Message field: Name of the event field that contains the message to send. If not specified, Cribl Edge sends a JSON representation of the whole event (regardless of whether **Send logs as** is set to JSON or plain text).

Source: Name of the source to send with logs. If you're sending logs as JSON objects (i.e., you've selected **Send logs as**: `application/json`), the event's `source` field (if set) will override this value.

Host: Name of the host to send with logs. If you're sending logs as JSON objects, the event's `host` field (if set) will override this value.

Service: Name of the service to send with logs. If you're sending logs as JSON objects, the event's `__service` field (if set) will override this value.

Datadog tags: List of tags to send with logs (e.g., `env:prod`, `env_staging:east`).

Severity: Default value for message severity. If you're sending logs as JSON objects, the event's `__severity` field (if set) will override this value. Defaults to `info`; the drop-down offers many other severity options.

Datadog uses the above five fields (`source`, `host`, `__service`, `tags`, and `__severity`) to enhance searches and UX.

Allow API key from events: If toggled to `Yes`, any API key in the `__agent_api_key` internal field will override the **API key** field's value. This option is useful if events originate from multiple Datadog Agent Sources, each configured with a different API key. (For further details, see Managing API Keys.)

Backpressure behavior: Specify whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`. Select `Gzip` to enable compression.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the

throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the

Backoff limit (ms). At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (s): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

If an event contains the internal field `__criblMetrics`, Cribl Edge will send it to Datadog as a metric event. Otherwise, Cribl Edge will send it as a log event.

You can use these fields to override outbound event values for log events:

- `__service`
- `__severity`

No internal fields are supported for metric events.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

For More Information

You might find these Datadog references helpful:

- Submit Metrics
- Send Logs
- Metrics Types

11.13. Exabeam Security Operations Platform

Cribl Edge supports sending data to the Exabeam security operations platform or (as some people think of it) SIEM.

The Exabeam Destination supersedes and improves upon the "old way" to get data from Cribl Edge to Exabeam using the Cribl Webhook Destination.

Type: Non-Streaming | TLS Support: Yes | PQ Support: No

Understanding How Exabeam and Cribl Edge Work Together

Exabeam parsers expect to receive events in original, unmodified form. To make sure this happens, you'll sometimes need to use a combination of Event Breakers and Pipelines. In general, knowing this can help you avoid problems or troubleshoot any that do crop up.

When you send data to Exabeam, Cribl Packs are your friends, because Cribl and Exabeam have partnered to create a set of Exabeam-specific Packs, each of which ensures that events from one common data source arrives at Exabeam in precisely the expected form. Cribl Packs also give you the option of dropping events on a per-data-source basis, so that Cribl Edge never sends irrelevant events to Exabeam. To view the Exabeam-specific Cribl Packs, navigate to the **Manage Packs** page's **New Pack** submenu, select **Add from Dispensary**, then enter `Exabeam` in the filter field.

Cribl strongly recommends that you use **only** the Exabeam-specific, Cribl-authored Packs with this Destination. Other Packs can perform transformations – especially reduction – that prevent the Exabeam parsers from triggering.

When the original data source is Microsoft Active Directory (AD) or LDAP, you'll need to set up an Exabeam Site Collector to perform the initial data capture.

Because Exabeam runs in the Google Cloud Storage (GCS) platform, some setting names include references to GCS.

Finally, Cribl recommends placing Cribl Edge Edge Nodes as close as possible to the sources of the events you want to send to Exabeam. This reduces the length of the path that data must traverse before Cribl Edge optimizes it with compression, reduction, dropping events, and/or redacting sensitive information.

Configuring Cloud Storage Permissions

For Cribl Edge to send data to Exabeam buckets, the following access permissions must be set on the Cloud Storage side:

- Fine-grained access control must be enabled on the buckets.
- The Google service account or user must have the Storage Admin or Owner role.

For details, see the Cloud Storage Overview of Access Control and Understanding Roles documentation.

Creating the Cribl Cloud Collector in Exabeam

Log into the Exabeam Security Operations Platform, and in the **Home** page, click **Cloud Collectors**. In the resulting page, click the **Cribl** tile to see a list of Cribl Collectors (the first time you do this, the list will be empty).

Click **New Collector**, and in the resulting **Collector for Cribl** drawer:

1. Enter a name for your new Cribl Cloud Collector.
2. Optionally, in **Advanced Settings**, configure the **Metadata** value **TIMEZONE**
3. If your desired site already exists, select it from the **SITE** drop-down.
    - If your desired site does not exist yet, skip the drop-down and click the **manage your sites** link underneath. This will take you to a separate UI where you will enter a value for the **Metadata** value **SITENAME**, and Exabeam will generate the associated metadata value **SITE ID**. When you have done this, Exabeam will take you back to the previous UI, and you must then select your newly-created site from the **SITE** drop-down.
4. Click **Install**.

A `Hooray!` confirmation modal will now replace the **Collector for Cribl** drawer. In this modal:

1. Click the copy icon to the right of the encoded **Connection String**. Save this string in a secure and handy location – it is required when configuring the Cribl Edge Exabeam Destination.
2. Click **Go to Overview** to exit the modal.

Although optional, configuring metadata is useful because it allows a single instance of the Exabeam SIEM to differentiate between data from multiple sources. This is relevant for large enterprises whose IT infrastructure comprises many sources whose IP addresses overlap.

Configuring Cribl Edge to Output to Cloud Storage Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click **Collect** (Edge only). Next, click **Add Destination** at right. From the resulting drawer's tiles or the **Destinations** left nav, select **Exabeam**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles, select **Exabeam**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Exabeam Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Click **Autofill with Exabeam Connection String** to automatically enter the values for the following **Exabeam Settings**.

- **Google Cloud Storage bucket**: Name of the destination bucket – in Exabeam's Cribl Cloud Collector this is called **GCS Bucket Name**. This value can be a constant, or a JavaScript expression that can be evaluated only at init time. For example, you can reference a Global Variable like this: `myBucket-${C.vars.myVar}`.
- **Google Cloud Storage bucket region**: In Exabeam's Cribl Cloud Collector this is called **GCS Bucket Region**. It is the GCS region where the bucket is located.
- **Collector instance ID**: In Exabeam's Cribl Cloud Collector this is called **Instance ID**.
- **Access key**: In Exabeam's Cribl Cloud Collector this is called **Access Key**.
- **Secret**: In Exabeam's Cribl Cloud Collector this is called **Secret Key**.
- **Site name**: In Exabeam's Cribl Cloud Collector this is called **SITENAME**.
- **Site ID**: In Exabeam's Cribl Cloud Collector this is called **SITE ID**.
- **Timezone offset**: In Exabeam's Cribl Cloud Collector this is called **TIMEZONE**.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports `c*` wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Backpressure behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300` seconds (5 minutes).

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to final output location. Defaults to `100`.

Cribl Edge will also close any files larger than 10 MB.

Remove empty staging dirs: When toggled on (the default), deletes empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove empty staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates (such as self-signed certificates, for example) that cannot be verified against a valid Certificate Authority. Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Troubleshooting

Verify that you are sending the original, unmodified events. Study the applicable Exabeam parsers to make sure you understand exactly what they're doing. If you need to modify events, make sure that when you do, the relevant parsers still trigger.

Nonspecific messages from Google Cloud of the form `Error: failed to close file` can indicate problems with the permissions listed above.

11.14. Filesystem/NFS

Filesystem is a non-streaming Destination type that Cribl Edge can use to output files to a local file system or a network-attached file system (NFS).

Type: Non-Streaming | TLS Support: N/A | PQ Support: N/A

In Cribl.Cloud, the Filesystem Destination is only available on hybrid, customer-managed Edge Nodes.

Configuring Cribl Edge to Output to Filesystem Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Filesystem**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Filesystem**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Filesystem definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Output location: Final destination for the output files.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (supported only on Linux, not Windows) exposes a **Parquet Settings** left tab, where you **must** configure certain options in order to export data in Parquet format.

Optional Settings

Staging location: Local filesystem location in which to buffer files before compressing and moving them to the final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Partitioning expression: JavaScript expression that defines how files are partitioned and organized. Default is date-based. If blank, Cribl Edge will fall back to the event's `__partition` field value (if present); or otherwise to the root directory of the **Output Location** and **Staging Location**.

Compress: Data compression format used before moving to final destination. Defaults to `gzip` (recommended). This setting is not available when **Data format** is set to `Parquet`.

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

To prevent files from being overwritten, Cribl appends a random sequence of 6 characters to the end of their names. File name prefix and suffix expressions do not bypass this behavior.

For example, if you set the **File name prefix expression** to `CriblExec` and set the **File name suffix expression** to `.csv`, the file name might display as `CriblExec-adPRWM.csv` with `adPRWM` appended.

Backpressure Behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other

options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported, and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: Set the target memory size for page segments. Generally, set lower values to improve reading speed, or set higher values to improve compression. Value must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files of this size will be closed and moved to final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to final output location. Defaults to `100`.

Cribl Edge will close files when **either** of the `Max file size (MB)` or the `Max file open time (sec)` conditions are met.

Header line: If set, this line will be written to the beginning of each output file, followed by a newline character. This can be useful for adding a header row to CSV files.

Add Output ID: When set to `Yes` (the default), adds the **Output ID** field's value to the staging location's file path. This ensures that each Destination's logs will write to its own bucket.

For a Destination originally configured in a Cribl Edge version below 2.4.0, the **Add Output ID** behavior will be switched **off** on the backend, regardless of this toggle's state. This is so that upon Cribl Edge upgrade and restart, any files pending in the original staging directory will not be lost. To enable this option for such Destinations, Cribl's recommended migration path is:

- Clone the Destination.

- Where Routes reference the original Destination, redirect them to instead reference the new, cloned Destination.

This way, the original Destination will process pending files (after an idle timeout), and the new, cloned Destination will process newly arriving events with **Add output ID** enabled.

Remove staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

To export events from an intermediate stage **within a Pipeline** to a file, see the Tee Function.

11.15. Honeycomb

Cribl Edge supports sending events to a Honeycomb dataset.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Honeycomb

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Honeycomb**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Honeycomb**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Honeycomb definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Dataset name: Name of the dataset to send events to. (E.g., `iLoveObservabilityDataset`.)

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays a field for you to enter the **API key** for the team to which the dataset belongs.
- **Secret**: This option exposes an **API key (text secret)** drop-down, in which you can select a stored secret that references the API key described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.

Status Code	Action
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular Destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between destination

cluster nodes.

11.16. InfluxDB

Cribl Edge supports sending data to InfluxDB (versions 1.x and 2.0.x) and InfluxDB Cloud.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output to InfluxDB

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **InfluxDB**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **InfluxDB**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this InfluxDB definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Write API URL: The URL of an InfluxDB cluster to send events to. (E.g., `http://localhost:8086/write`.)

Use v2 API: You can enable the InfluxDB v2 API with InfluxDB version 1.8 or later. This toggle defaults to `No` – which falls back to the v1 API, and displays a **Database** field.

If you toggle **Use v2 API** to `Yes`, Cribl Edge communicates using InfluxDB's v2 API, and instead displays these two fields:

- **Bucket**: Enter the bucket to write to. (Required.)
- **Organization**: The Organization ID corresponding to the specified **Bucket**. (Required in this configuration, although InfluxDB v.1.8 will ignore it.)

Database: Name of the database on which to write data points. (Required.)

Optional Settings

Timestamp precision: Sets the precision for the supplied UNIX time values. Defaults to `Milliseconds`.

Dynamic value fields: When enabled, Cribl Edge will pull the value field from the metric name. (E.g., `db.query.user` will use `db.query` as the measurement and `user` as the value field). Defaults to `Yes`.

Value field name: Name of the field in which to store the metric when sending to InfluxDB. This will be used as a fallback if dynamic name generation is enabled but fails. Defaults to `value`.

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Authentication

Use the **Authentication type** drop-down to select one of these options:

None: This default setting does not use authentication.

Auth token: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.

Auth token (text secret): This option exposes a **Token (text secret)** drop-down, in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.

Basic: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.

Basic (credentials secret): This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are **not** authorized by a trusted CA (e.g., the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

11.17. MinIO

MinIO is a non-streaming Destination type, to which Cribl Edge can output objects.

Type: Non-Streaming | TLS Support: Configurable | PQ Support: No

Configuring Cribl Edge to Output to MinIO Destinations

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **MinIO**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **MinIO**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this MinIO definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

MinIO endpoint: MinIO service URL (e.g., http://minioHost:9000).

MinIO bucket name:Name of the destination MinIO bucket. This value can be a constant, or a JavaScript expression that will be evaluated only at init time. E.g., referencing a Global Variable: `myBucket-${C.vars.myVar}`. Ensure that the bucket already exists, otherwise MinIO will generate "bucket does not exist" errors.

Event-level variables are not available for JavaScript expressions. This is because the bucket name is evaluated only at Destination initialization. If you want to use event-level variables in file paths, Cribl recommends specifying them in the **Partitioning Expression** field (described below), because this is evaluated for each file.

Staging location: Filesystem location in which to locally buffer files before compressing and moving to final destination. Cribl recommends that this location be stable and high-performance.

The **Staging location** field is not displayed or available on Cribl.Cloud-managed Edge Nodes.

Key prefix: Root directory to prepend to path before uploading. Enter a constant, or a JS expression enclosed in single quotes, double quotes, or backticks.

Prefix to apply to files/objects before uploading to the specified bucket. MinIO will display key prefixes as folders.

Data format: The output data format defaults to `JSON`. `Raw` and `Parquet` are also available. Selecting `Parquet` (supported only on Linux, not Windows) exposes a Parquet Settings left tab, where you **must** configure certain options in order to export data in Parquet format.

Optional Settings

Partitioning expression: JavaScript expression that defines how files are partitioned and organized. Default is date-based. If blank, Cribl Edge will fall back to the event's `__partition` field value (if present); or otherwise to the root directory of the **Output Location** and **Staging Location**.

Cribl Edge's internal `__partition` field can be populated in multiple ways. The precedence order is: explicit **Partitioning expression** value -> `${host}/${sourcetype}` (default) **Partitioning expression** value -> user-defined `event.__partition`, set with an Eval Function (takes effect only where this **Partitioning expression** field is blank).

Compress: Data compression format used before moving to final destination. Defaults to `gzip` (recommended). This setting is not available when **Data format** is set to `Parquet`.

File name prefix expression: The output filename prefix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `CriblOut`.

File name suffix expression: The output filename suffix. Must be a JavaScript expression (which can evaluate to a constant), enclosed in quotes or backticks. Defaults to `` `.${C.env["CRIBL_WORKER_ID"]}.${__format}${__compression === "gzip" ? ".gz" : ""}` ``, where `__format` can be `json` or `raw`, and `__compression` can be `none` or `gzip`.

Backpressure behavior: Select whether to block or drop events when all receivers are exerting backpressure. (Causes might include an accumulation of too many files needing to be closed.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

How MinIO Composes File Names

The full path to a file consists of:

```
<bucket_name>/<keyprefix><partition_expression | __partition><file_name_prefix>
<filename>.<extension>
```

As an example, assume that the **MinIO bucket name** is `bucket1`, the **Key prefix** is `aws`, the **Partitioning expression** is `` `${host}/${sourcetype}` ``, the source is undefined, the **File name prefix** is the default `CriblOut`, and the **Data format** is `json`. Here, the full path as displayed in MinIO would have this form: `/bucket1/aws/192.168.1.241/undefined/CriblOut-<randomstring>0.json`

Although MinIO will display the **Key prefix** and **Partitioning expression** values as folders, both are actually just part of the overall key name, along with the file name.

Authentication

Use the **Authentication Method** drop-down to select one of these options:

Auto: This default option uses the AWS instance's metadata service to automatically obtain short-lived credentials from the IAM role attached to an EC2 instance, local credentials, sidecar, or other source. The attached IAM role grants Cribl Edge Workers access to authorized AWS resources. Can also use the environment variables `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY`. Works only when running on AWS.

Manual: If not running on AWS, you can select this option to enter a static set of user-associated IAM credentials (your access key and secret key) directly or by reference. This is useful for Workers not in an AWS VPC, e.g., those running a private cloud. The **Manual** option exposes these corresponding additional fields:

- **Access key**: Enter your AWS access key. If not present, will fall back to the `env.AWS_ACCESS_KEY_ID` environment variable, or to the metadata endpoint for IAM role credentials.
- **Secret key**: Enter your AWS secret key. If not present, will fall back to the `env.AWS_SECRET_ACCESS_KEY` environment variable, or to the metadata endpoint for IAM credentials.

The values for **Access key** and **Secret key** can be a constant, or a JavaScript expression (such as `${C.env.MY_VAR}`) enclosed in quotes or backticks, which allows configuration with environment variables.

Secret: If not running on AWS, you can select this option to supply a stored secret that references an AWS access key and secret key. This option exposes a **Secret key pair** drop-down, in which you can select a

stored secret that references the set of user-associated IAM credentials described above. A **Create** link is available to store a new, reusable secret.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Parquet Settings

To write out Parquet files, note that:

- On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.
- Cribl Edge Workers support Parquet only when running on Linux, not on Windows.
- See Working with Parquet for pointers on how to avoid problems such as data mismatches.

Automatic schema: Toggle on to automatically generate a Parquet schema based on the events of each Parquet file that Cribl Edge writes. When toggled off (the default), exposes the following additional field:

- **Parquet schema**: Select a schema from the drop-down.

If you need to modify a schema or add a new one, follow the instructions in our Parquet Schemas topic. These steps will propagate the freshest schema back to this drop-down.

Parquet version: Determines which data types are supported, and how they are represented. Defaults to `2.6`; `2.4` and `1.0` are also available.

Data page version: Serialization format for data pages. Defaults to `V2`. If your toolchain includes a Parquet reader that does not support `V2`, use `V1`.

Group row limit: The number of rows that every group will contain. The final group can contain a smaller number of rows. Defaults to `10000`.

Page size: Set the target memory size for page segments. Generally, set lower values to improve reading speed, or set higher values to improve compression. Value must be a positive integer smaller than the **Row group size** value, with appropriate units. Defaults to `1 MB`.

Log invalid rows: Toggle to `Yes` to output up to 20 unique rows that were skipped due to data format mismatch. Log level must be set to `debug` for output to be visible.

Write statistics: Leave toggled on (the default) if you have Parquet tools configured to view statistics – these profile an entire file in terms of minimum/maximum values within data, numbers of nulls, etc.

Write page indexes: Leave toggled on (the default) if your Parquet reader uses statistics from Page Indexes to enable page skipping. One Page Index contains statistics for one data page.

Write page checksum: Toggle on if you have configured Parquet tools to verify data integrity using the checksums of Parquet pages.

Metadata (optional): The metadata of files the Destination writes will include the properties you add here as key-value pairs. For example, one way to tag events as belonging to the OCSF category for security findings would be to set **Key** to `OCSF Event Class` and **Value** to `2001`.

Advanced Settings

Max file size (MB): Maximum uncompressed output file size. Files of this size will be closed and moved to final output location. Defaults to `32`.

Max file open time (sec): Maximum amount of time to write to a file. Files open for longer than this limit will be closed and moved to final output location. Defaults to `300`.

Max file idle time (sec): Maximum amount of time to keep inactive files open. Files open for longer than this limit will be closed and moved to final output location. Defaults to `30`.

Max open files: Maximum number of files to keep open concurrently. When exceeded, the oldest open files will be closed and moved to final output location. Defaults to `100`.

Cribl Edge will close files when **either** of the `Max file size (MB)` or the `Max file open time (sec)` conditions is met.

Max concurrent file parts: Maximum number of parts to upload in parallel per file. A value of `1` tells the Destination to send the whole file at once. When set to `2` or above, IAM permissions must include those required for multipart uploads. Defaults to `4`; highest allowed value is `10`.

Add Output ID: When set to `Yes` (the default), adds the **Output ID** field's value to the staging location's file path. This ensures that each Destination's logs will write to its own bucket.

> For a Destination originally configured in a Cribl Edge version below 2.4.0, the **Add Output ID** behavior will be switched **off** on the backend, regardless of this toggle's state. This is to avoid losing any files pending in the original staging directory, upon Cribl Edge upgrade and restart. To enable this option for such Destinations, Cribl's recommended migration path is:
>
> - Clone the Destination.
> - Redirect the Routes referencing the original Destination to instead reference the new, cloned Destination.
>
> This way, the original Destination will process pending files (after an idle timeout), and the new, cloned Destination will process newly arriving events with **Add output ID** enabled.

Remove staging dirs: Toggle to `Yes` to delete empty staging directories after moving files. This prevents the proliferation of orphaned empty directories. When enabled, exposes this additional option:

- **Staging cleanup period**: How often (in seconds) to delete empty directories when **Remove staging dirs** is enabled. Defaults to `300` seconds (every 5 minutes). Minimum configurable interval is `10` seconds; maximum is `86400` seconds (every 24 hours).

Region: Region where the MinIO service/cluster is located. Leave blank when using a containerized MinIO.

Object ACL: ACL (Access Control List) to assign to uploaded objects. Defaults to `Private`.

Storage class: Select a storage class for uploaded objects. Defaults to `Standard`.

Server-side encryption: Server side encryption type for uploaded objects. Defaults to `none`.

Signature version: Signature version to use for signing MinIO requests. Defaults to `v4`.

Reuse connections: Whether to reuse connections between requests. The default setting (`Yes`) can improve performance.

Reject unauthorized certificates: Whether to accept certificates that cannot be verified against a valid Certificate Authority (e.g., self-signed certificates). Defaults to `Yes`.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

IAM Permissions

The following permissions are always needed to write to an Amazon S3-compatible object store:

- `s3:ListBucket`
- `s3:GetBucketLocation`
- `s3:PutObject`

If your Destination needs to do multipart uploads to S3, two more permissions are needed:

- `kms:GenerateDataKey`
- `kms:Decrypt`

See the AWS documentation.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Field for this Destination:

- `__partition`

Troubleshooting

See also AWS Sources/Destinations & S3-Compatible Stores for information on common errors.

11.18. NetFlow

Cribl Edge supports receiving NetFlow v5 data via UDP.

Type: **Push** | TLS Support: **No** | Event Breaker Support: **No**

This Source ingests NetFlow records similarly to how it ingests events from other upstream senders: fields are broken out, and the message header is included with each record. If you prefer to render NetFlow data as metrics, use a pre-processing Pipeline or a Route.

Configuring Cribl Edge to Receive NetFlow Data

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **+ Add Source** at left. From the resulting drawer's tiles, select [**Push** >] **NetFlow**. Next, click either **+ Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Sources** (Stream) or **More** > **Sources** (Edge). From the resulting page's tiles or left nav, select [**Push** >] **NetFlow**. Next, click **Add Source** to open a **New Source** modal that provides the options below.

Sending large numbers of UDP events per second can cause Cribl.Cloud to drop some of the data. This results from restrictions of the UDP protocol.

To minimize the risk of data loss, deploy a hybrid Stream Worker Group with customer-managed Worker Nodes as close as possible to the UDP sender. Cribl also recommends tuning the OS UDP buffer size.

General Settings

Input ID: Enter a unique name to identify this NetFlow Source definition.

Address: Enter the hostname/IP to listen for NetFlow data. For example: `localhost`, `0.0.0.0`, or `::`.

Port: Enter the port number.

Tags: Optionally, add tags that you can use to filter and group Sources in Cribl Edge's **Manage Sources** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

In this section, you can optionally specify persistent queue storage, using the following controls. This will buffer and preserve incoming events when a downstream Destination is down, or exhibiting backpressure.

> On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the **Enable Persistent Queue** toggle. If enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB disk space allocated per PQ-enabled Source, per Worker Process.
>
> The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue. This limit is not configurable. For configurable queue size, compression, mode, and other options below, use a hybrid Group.
>
> For more about PQ modes, see Always On versus Smart Mode.

Enable Persistent Queue: Defaults to `No`. When toggled to `Yes`:

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Cribl Edge has read them. Defaults to `42`.

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Source will stop queueing data and block incoming data. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this field's specified path, Cribl Edge will append `/<worker-id>/inputs/<input-id>`.

Compression: Optional codec to compress the persisted data after a file is closed. Defaults to `None`; `Gzip` is also available.

You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Fields

In this section, you can add Fields to each event using Eval-like functionality.

- **Name**: Field name.
- **Value**: JavaScript expression to compute field's value, enclosed in quotes or backticks. (Can evaluate to a constant.)

Pre-Processing

In this section's **Pipeline** drop-down list, you can select a single existing Pipeline to process data from this input before the data is sent through the Routes.

Advanced Settings

IP allowlist regex: Regex matching IP addresses that are allowed to establish a connection. Defaults to `.*` (that is, all IPs).

IP denylist regex: Regex matching IP addresses whose messages you want this Source to ignore. Defaults to `^$` (that is, every specific IP address in the list). This takes precedence over the allowlist.

UDP socket buffer size (bytes): Optionally, set the `SO_RCVBUF` socket option for the UDP socket. This value tells the operating system how many bytes can be buffered in the kernel before events are dropped. Leave blank to use the OS default. Minimum: `256`. Maximum: `4294967295`.

It may also be necessary to increase the size of the buffer available to the `SO_RCVBUF` socket option. Consult the documentation for your operating system for a specific procedure.

Setting this value will affect OS memory utilization.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Connected Destinations

Select **Send to Routes** to enable conditional routing, filtering, and cloning of this Source's data via the Routing table.

Select **QuickConnect** to send this Source's data to one or more Destinations via independent, direct connections.

What Fields to Expect

The NetFlow Source does minimal processing of the incoming UDP messages to remain consistent with the internal Cribl event model. For each UDP message received on the socket, you can expect an event with the following fields. We have organized these fields into categories to make them easier to grasp; the categories are our own, and not from the NetFlow specifications. The field definitions are mostly copied from Cisco NetFlow documentation.

General

- `_time`: The UNIX timestamp (in seconds) at which the message was received by Cribl Edge.
- `source`: A string in the form `udp|<remote IP address>|<remote port>`, indicating the remote sender.
- `host`: The hostname of the machine running Cribl Edge that ingested this event.
- `inputInt`: SNMP index of input interface; always set to zero.
- `outputInt`: SNMP index of output interface.
- `protocol`: IP protocol type (for example, TCP = 6; UDP = 17) of the observed network flow.
- `tcpFlags`: Cumulative logical OR of TCP flags in the observed network flow.
- `tos`: IP type of service; switch sets it to the ToS of the first packet of the flow.
- `icmpType`: Type of the ICMP message. Only present if the protocol is ICMP.
- `icmpCode`: Code of the ICMP message. Only present if the protocol is ICMP.

Because this Source reads input data directly from bytes in a compact format, there is nothing suitable to put in a `_raw` field, and the Source does not add a `_raw` field to events.

Flow Source and Destination

- `srcAddr`: Source IP address; in case of destination-only flows, set to zero.
- `dstAddr`: Destination IP address.

- `nextHop`: IP address of next hop router.
- `srcPort`: TCP/UDP source port number.
- `dstPort`: TCP/UDP destination port number. If this is an ICMP flow, this field is a combination of ICMP type and ICMP code, which are broken out separately as `icmpType` and `icmpCode` fields.
- `srcMask`: Source address prefix mask bits.
- `dstMask`: Destination address prefix mask bits.
- `srcAs`: Autonomous system number of the source, either origin or peer.
- `dstAs`: Autonomous system number of the destination, either origin or peer.

Flow Statistics

- `packets`: Packets in the flow.
- `octets`: Total number of Layer 3 bytes in the packets of the flow.
- `startTime`: System uptime, in milliseconds, at the start of the flow.
- `endTime`: System uptime, in milliseconds, at the time the last packet of the flow was received.
- `durationMs`: `endTime` minus `startTime`.

Header

The following are subfields within the `header` field:

- `count`: Number of flows exported in this flow frame (protocol data unit, or PDU).
- `sysUptimeMs`: Current time in milliseconds since the export device booted
- `unixSecs`: Current seconds since 0000 UTC 1970.
- `unixNsecs`: Residual nanoseconds since 0000 UTC 1970.
- `flowSequence`: Sequence counter of total flows seen.
- `engineType`: Type of flow switching engine.
- `engineId`: ID number of the flow switching engine.
- `samplingMode`: The first two bits of what Cisco NetFlow documentation calls `SAMPLING_INTERVAL`. These bits specify a sampling mode.
- `samplingInterval`: The remaining 14 bits of what Cisco NetFlow documentation calls `SAMPLING_INTERVAL`. These bits specify a sampling interval.
- `version`: NetFlow export format version number.

Also, the internal fields listed below will be present.

Internal Fields

Cribl Edge uses a set of internal fields to assist in handling of data. These "meta" fields are **not** part of an event, but they are accessible, and functions can use them to make processing decisions.

Fields accessible for this Source:

- `__bytes`
- `__inputId`
- `_time`

UDP Tuning

Incoming UDP traffic is put into a buffer by the Linux kernel. Cribl will drain from that buffer as resources are available. At lower throughput levels, and with plenty of available processes, this isn't an issue. As you scale up, however, the default size of that buffer may be too small.

You can check the current buffer size with:

```
$ sysctl net.core.rmem_max
```

A typical value of about 200 KB is far too small for an even moderately busy syslog server. You can check the health of UDP with the following command. Check the `packet receive errors` line.

```
$ netstat -su
```

If `packet receive errors` is more than zero, you have lost events, which is a particularly serious problem if the number of errors is increasing rapidly. This means you need to increase your `net.core.rmem_max` setting (see earlier).

You can update the live settings, but you'll also need to change the boot-time setting so next time you reboot everything is ready to roll.

Live change, setting to 25 MB:

```
$ sysctl -w net.core.rmem_max=26214400
net.core.rmem_max = 26214400
$ sysctl -w net.core.rmem_default=26214400
net.core.rmem_default = 26214400
```

For the permanent settings change, add the following lines to `/etc/sysctl.conf`:

```
net.core.rmem_max=26214400
net.core.rmem_default=26214400
```

We recommend you track a few things related to UDP receiving:

- The `netstat -su` command, watching for errors.
- The **Status** tab in the UDP (Raw) Source. In particular, watch for dropped messages. They could indicate you need a bigger buffer under **Advanced Settings** (default: 1000 events). They could also indicate your Worker is encountering pressure further down the Pipeline.
- Especially if you increase your kernel receive buffer as above, watch your Worker processes' memory usage.

11.19. OpenTelemetry (OTel)

Cribl Edge supports sending events to OpenTelemetry Protocol (OTLP)-compliant targets. (Cribl Edge can receive OTel events through the OTel Source.) Besides native OTel Trace and Metric events, you can send Cribl Edge's Gauge metric events through this OTel Destination.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Protocol and Transport Support

In Cribl Edge 4.1 and later, this Destination supports either of the transports that OTLP specifies: gRPC or HTTP. OTLP defines Protocol buffer (Protobuf) schemas for its payloads (requests and responses). With the HTTP transport, this Destination supports Binary Protobuf payload encoding, but currently does not support JSON Protobuf.

When configuring Pipelines (including pre-processing and post-processing Pipelines), you need to ensure that events sent to this Destination conform to the relevant Protobuf specification:

- For traces, opentelemetry-proto/trace.proto at v0.9.0 · open-telemetry/opentelemetry-proto
- For metrics, opentelemetry-proto/metrics.proto at v0.9.0 · open-telemetry/opentelemetry-proto

The OTel Destination will drop non-conforming events. If the Destination encounters a parsing error when trying to convert an event to OTLP, it discards the event and Cribl Edge logs the error.

Configuring Cribl Edge to Output to OTel

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **OpenTelemetry**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **OpenTelemetry**.
Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this OTel output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Endpoint: Where to send events, in any of a variety of formats (FQDN, PQDN, IP address and port, etc). Supports both IPv4 and IPv6 – IPv6 addresses must be enclosed in square brackets. The same endpoint is used for both Traces and Metrics. If no port is specified, we normally default to the standard port for OTel Collectors, 4137 – however, if TLS is enabled or the endpoint is an HTTPS-based URL, we default instead to port 443.

To proxy outbound HTTP/S requests, see System Proxy Configuration.

Optional Settings

Protocol: Use the drop-down to choose the protocol to use when sending data: `gRPC` (the default), or `HTTP`. When set to `HTTP`, the UI add the Retries section (left tab) and displays two additional settings:

- **Traces endpoint override**: By default, the Destination will send traces to `<endpoint>/v1/traces`, where `<endpoint>` is what you specified for the **Endpoint** setting above. To send traces to a different endpoint, enter that endpoint here.
- **Metrics endpoint override**: By default, the Destination will send traces to `<endpoint>/v1/metrics`, where `<endpoint>` is what you specified for the **Endpoint** setting above. To send metrics to a different endpoint, enter that endpoint here.

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Client Side)

TLS is available only when **General Settings** > **Protocol** is set to `gRPC`.

Enabled Defaults to `No`. When toggled to `Yes`:

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Enter the bearer token that must be included in the authorization header. Since OpenTelemetry runs over gRPC, authorization headers are sent as Metadata entries which are essentially key-value pairs. E.g.: `Bearer <your-configured-token>`.
- **Auth token (text secret)**: This option exposes a drop-down in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.
- **Basic**: This default option displays fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, queueing is stopped and data blocking is applied. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. This will be of the form: `your/path/here/<worker-id>/<output-id>`. Defaults to: `$CRIBL_HOME/state/queues`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`. `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output – both metric events, as dimensions; and log events, as labels. Supports wildcards.

By default, includes `cribl_pipe` (Cribl Edge Pipeline that processed the event).

Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

This tab is displayed when **General Settings** > **Optional Settings** > **Protocol** is set to `HTTP`.

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

This tab's options depend on whether **General Settings** > **Protocol** is set to the `gRPC` or `HTTP` transport. The common settings directly are displayed for both transports.

Common Settings

Compression: Compression type to apply to messages sent to the OpenTelemetry endpoint. `Gzip` (the default) and `None` are available for both protocols; `Deflate` is available for gRPC only.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30` sec.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes

to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` sec.

Environment: Optionally, specify a single Git branch on which to enable this configuration. If this field is empty, the config will be enabled everywhere.

Additional Settings for gRPC

When **General Settings** > **Protocol** is set to `gRPC`, the **Advanced Settings** tab adds the following options.

Connection timeout: Amount of time (milliseconds) to wait for the connection to establish before retrying. Defaults to `10000` (10 sec.).

Keep alive time (seconds): How often the sender should ping the peer to keep the connection alive. Defaults to `30`.

Metadata: Extra information to send with each gRPC request. Click **Add Metadata** to add each item as a **Key**-**Value** pair. The **Key** field is arbitrary. The **Value** field is a JavaScript expression that is evaluated just once, when this Destination is initialized. If you pass credentials as metadata, Cribl recommends using C.Secret().

Additional Settings for HTTP

When **General Settings** > **Protocol** is set to `HTTP`, the **Advanced Settings** tab adds the following options.

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Keep alive: By default, Cribl Edge sends `Keep-Alive` headers to the remote server and preserves the connection from the client side up to a maximum of 120 seconds. Toggle this off if you want Cribl Edge to close the connection immediately after sending a request.

Extra HTTP headers: Click **Add Header** to define additional HTTP headers to pass to all events. Each row is a **Name**-**Value** pair. Values will be sent encrypted. You can also add headers dynamicall,y on a per-event basis, in the `__headers` field.

Safe headers: Add headers here to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

11.20. SentinelOne DataSet

Cribl Edge can send log events to the SentinelOne/Scalyr DataSet platform via the DataSet API. This Destination sends batches of events, as JSON, to that API's addEvents endpoint.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to DataSet

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **DataSet**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **DataSet**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition.

Authentication Settings

Use the **Authentication method** buttons to select one of these options:

- **Manual**: Displays a field for you to enter an **API key** that is available in your DataSet profile.
- **Secret**: This option exposes an **API key (text secret)** drop-down to select a stored secret that references an API key. A **Create** link is available to store a new, reusable secret.

API key: Enter your DataSet API key that has `Log Write Access`.

Optional Settings

DataSet site: Select the US (default), `Europe`, or `Custom` region. If you select `Custom`, enter your custom endpoint URL.

Message field: Name of the event field that contains the message to send. If not specified, Cribl Edge sends all non-internal fields of events passing through the Destination. If specified, we follow this logic:

- If an event does not contain the specified field, send the whole event (except internal fields).
- If an event has the specified field, and the field's value is a non-object, send the event in the format: `{ message: <value from event> }`.
- If an event has the specified field, and the field's value is an object, send the event in the format: `{ <all fields from the object> }`.

Exclude fields: Fields to exclude from the event if the **Message field** either is unspecified or refers to an object. Ignored if the **Message field** is a string, number, or boolean. If empty, Cribl Edge sends all non-internal fields.

Default exclude fields are `sev`, `_time`, `ts`, and `thread`. We automatically send these fields as metadata of the event, in DataSet's required format. This is to avoid charges for field bytes – metadata bytes do not count toward ingestion.

Server/host field: Name of the event field that contains the server or host that generated the event. Cribl Edge groups events by the value of this field, and gives them a unique session token to conform to the DataSet API. Each group is sent out as a separate batch; therefore, Cribl recommends specifying a field with a low cardinality, to avoid queuing up many different batches at the Destination. If not specified, or not a string, the implicit default value is `cribl_<outputId>`.

Timestamp field: Name of the event field that contains the event timestamp. Cribl Edge sends this value as part of each event's metadata, not as an attribute field on the event.

Timestamps are automatically converted to a nanosecond-precision string. If an event does not contain the field specified **Timestamp field**, or if the value cannot be converted into a nanosecond-precision string, Cribl Edge assigns a timestamp using the first valid output returned from `ts`, `_time`, or `Date.now()`, in that order.

Severity: Use the drop-down to assign a default value to the `severity` field (which is sent as event metadata, not as an attribute field). Cribl Edge falls back to this value when an event contains no valid `sev` or `__severity` field. DataSet's severity model ranges from `0` least-severe (finest) to `6` most-severe (fatal).

- Where an event's `sev` field contains an integer in this range, Cribl Edge passes it through as the severity.
- Where the `sev` field contains a string matching DataSet's enum (`finest`, `finer`, `fine`, `info`, `warning`, `error`, `fatal`), Cribl Edge converts it to the corresponding integer.

Backpressure behavior: Specify whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This section is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space the queue is allowed to consume. Once this limit is reached, Cribl Edge stops queueing and applies the fallback **Queue-full behavior**. Enter a numeral with units of KB, MB, etc.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`. Select `Gzip` to enable compression.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear persistent queue: Click this button if you want to flush out files that are currently queued for delivery to this Destination. A confirmation modal will appear. (Appears only after **Output ID** has been defined.)

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, `429 (Too Many Requests)` is the only response code configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.

- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Defaults to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests before blocking. This is set per Worker Process. Defaults to `5`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Click **Add header** to define **Name**/**Value** pairs to pass as additional HTTP headers.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers here to declare them as safe to log in plaintext. (Sensitive headers like `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Internal Fields

The `__severity` field is included in the severity assignment order, after the `sev` field. The order is `sev`, `__severity`, then the configured default **Severity**.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

11.21. SignalFx

Cribl Edge supports sending events to SignalFx.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to SignalFx

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **SignalFx**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **SignalFx**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this SignalFx definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Realm: SignalFx realm name (e.g., `us0`). Required.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays an **Auth token** field for you to enter your SignalFx API access token. See SignalFx's Manage Tokens topic.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the API access token described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.

Status Code	Action
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` second.

Extra HTTP headers: Click **Add Header** to insert extra headers as **Name**/**Value** pairs. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

Notes About SignalFx

For details on integrating with SignalFx, see the SignalFx Developers Guide. Especially relevant is the SignalFx HTTP Send Metrics Reference.

11.22. SNMP Trap

Cribl Edge supports forwarding of SNMP Traps out.

Type: Streaming | TLS Support: No | PQ Support: No

Configuring Cribl Edge to Forward to SNMP Traps

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **SNMP Trap**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **SNMP Trap**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this SNMP Trap definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

SNMP trap destinations: Each row provides the following fields:

- **Address**: Destination host.
- **Port**: Destination port. Defaults to `162`.

Add Destination: Click to specify additional SNMP trap destinations to forward traps to on new rows.

Optional Settings

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Considerations for Working with SNMP Traps Data

- It's possible to work with SNMP metadata (i.e., we'll decode the packet). Options include dropping, routing, etc. However, packets **cannot** be modified and sent to another SNMP Destination.
- SNMP packets can be forwarded to non-SNMP Destinations (e.g., Splunk, Syslog, S3, etc.).
- SNMP packets can be forwarded to other SNMP Destinations. However, the contents of the incoming packet cannot be modified – i.e., we'll forward the packets verbatim as they came in.
- Non-SNMP input data **cannot** be sent to SNMP Destinations.

11.23. Sumo Logic

Cribl Edge can send logs and metrics to Sumo Logic over HTTP. Sumo Logic offers a Hosted Collector that supports HTTP Sources to receive data over an HTTP POST request.

To send traces to Sumo Logic you can use Cribl's OpenTelemetry Destination pointing to Sumo Logic's OTLP/HTTP Source.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

How Sumo Logic Handles Data

- When an event contains the internal field `__criblMetrics`, it's sent to Sumo Logic as a metric event. Otherwise, it's sent as a log event.
- Data sent to Sumo Logic needs to be UTF-8 encoded and they recommend a data payload have a size, before compression, of 100 KB to 1 MB.
- Sumo Logic may impose throttling and caps on your log ingestion to prevent your account from using On-Demand Capacity, see Sumo Logic's manage ingestion documentation for details.

Prerequisites

In Sumo Logic, create or retrieve an HTTP Logs and Metrics Source's unique endpoint URL. See Sumo Logic's HTTP Logs and Metrics Source documentation for details. You will need the `Manage Collectors` role capability in Sumo Logic.

After creating the Source in Sumo Logic, the URL associated with the Source is displayed. Copy the endpoint URL so you can enter it when you configure the Cribl Edge Sumo Logic Destination.

Configure a Sumo Logic Destination

In Cribl Edge, set up a Sumo Logic Destination.

1. From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:
    - To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Sumo Logic**. Next, click either **Add Destination** or (if displayed) **Select Existing**.

- Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Sumo Logic**. Next, click **Add Destination** to open a **New Destination** modal.

2. In the Destination modal, configure the following under **General Settings**:
   - **Output ID**: Enter a unique name to identify this Sumo Logic Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.
   - **API URL**: Enter the endpoint URL of the Sumo Logic HTTP Source. This is provided by Sumo Logic after creating the Source, see prerequisites for details. For example, `https://endpoint6.collection.us2.sumologic.com/receiver/v1/http/<long-hash>`.
3. Next, you can configure the following **Optional Settings**:
   - **Custom source name** and **Custom source category** are unique settings to Sumo Logic. These allow you to override the Sumo Logic HTTP Source's **Source Name** and **Source Category** values with HTTP headers. Alternatively, you can define the `__sourceName` and `__sourceCategory` fields on events to assign a custom value at the event level.

   The remaining configurations are Cribl settings that you'll find across many Cribl Destinations.
   - **Backpressure behavior**: Whether to block, drop, or queue events when all receivers are exerting backpressure.
   - **Tags**: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.
   - **Data Format**: This drop-down defaults to `JSON`. Change this to `Raw` if you prefer to preserve outbound events' raw format instead of JSONifying them.
4. Optionally, configure any Persistent Queue, Processing, Retries, and Advanced settings outlined in the below sections.
5. Click **Save**, then **Commit & Deploy**.
6. Verify that data is searchable in Sumo Logic. See the Verify Data Flow section below.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

> This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `1024` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass as additional HTTP headers. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

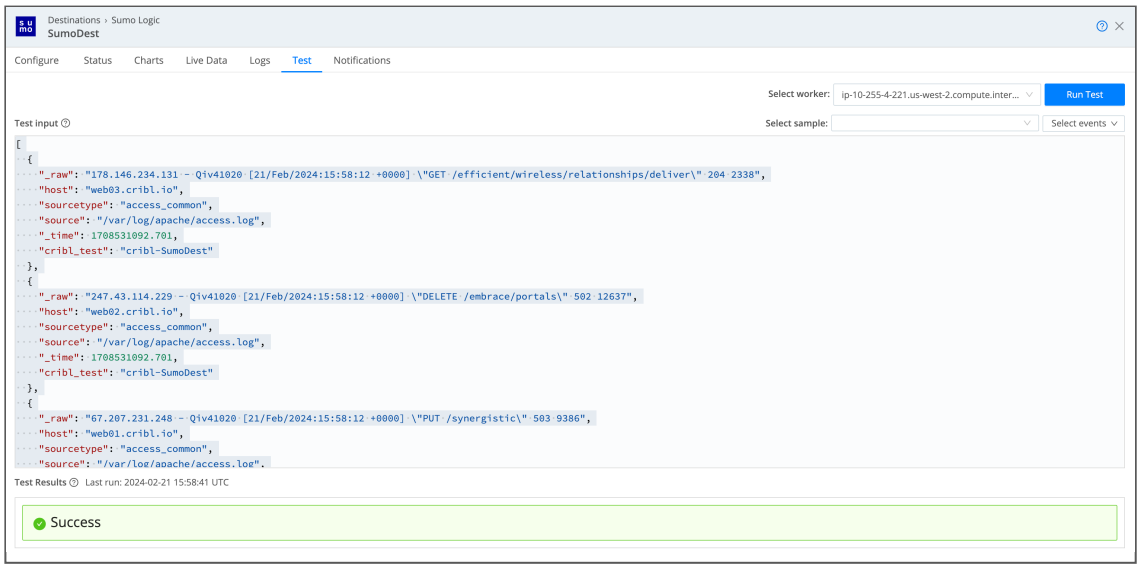
Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Verify Data Flow

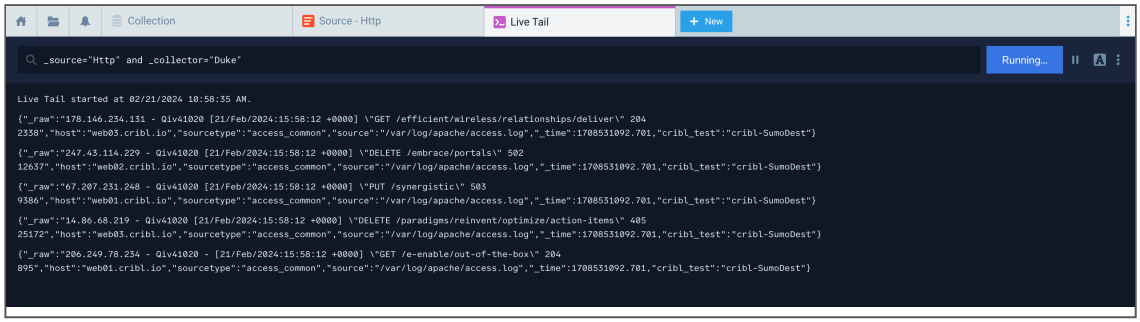
To verify data flow, we'll use the Destination's **Test** feature while running a **Live Tail** session in Sumo Logic.

1. In Sumo Logic, run a new Live Tail session against the name of the HTTP Logs and Metrics Source or the **Custom source name** you defined when you configured the Sumo Logic Destination. You'll need to use the `_source` metadata field. For example, if the name of the Source is `HTTP` your search would be `_source="HTTP"`. Ensure you've clicked the **Run** button to start the Live Tail session.
2. In Cribl Edge, open the Destination configuration modal and select the **Test** tab. You can leave the default data or select from the available samples. Click **Run Test**.



Example Destination Test

3. Look back to your Sumo Logic Live Tail session and you should see the sample data from the Cribl test displayed in your Live Tail results.



Example Live Tail Results

Troubleshooting

If you receive an error when verifying data flow, take note of the response code and reference Sumo Logic's documentation for common issues to investigate.

Notes on HTTP-Based Outputs

- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, the connection will be thrown away, and a new one will be reattempted. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), a new connection will be established for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable Round-robin DNS to better balance distribution of events between destination cluster nodes.

11.24. Syslog

Cribl Edge supports sending out data over syslog via TCP or UDP.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

This Syslog Destination supports RFC 3164 and RFC 5424. Before you configure this Destination, review Understanding Syslog Format Options below, to ensure that your configuration will structure compliant outbound events that downstream services can read.

Configuring Cribl Edge to Output Data in Syslog Format

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Syslog**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Syslog**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Syslog definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Protocol: The network protocol to use for sending out syslog messages. Defaults to `TCP`; `UDP` is also available.

Load balancing: This option is displayed only when the **Protocol** is set to `TCP`. When enabled (default), lets you specify multiple destinations. See Load Balancing Settings below. With the default `No` setting, if you notice that Cribl Edge is not sending data to all possible IP addresses, enable **Advanced Settings** > **Round-robin DNS**.

The following two fields appear **only** with the `No` setting.

Address: Address/hostname of the receiver.

Port: Port number to connect to on the host.

Max record size: Displayed when **Protocol** is set to `UDP`. Sets the maximum size of syslog messages. Defaults to `1500`, and must be ≤ `2048`. To avoid packet fragmentation, keep this value <= the MTU (maximum transmission unit for the network path to the Destination system).

Optional Settings

Facility: Default value for message facility. If set, will be overwritten by the value of `__facility`. Defaults to `user`.

Severity: Default value for message severity. If set, will be overwritten by the value of `__severity`. Defaults to `notice`.

App name: Default value for application name. If set, will be overwritten by the value of `__appname`. Defaults to `Cribl`.

Message format: The syslog message format supported by the receiver. Defaults to `RFC3164`.

Timestamp format: The timestamp format to use when serializing an event's time field. Options include `Syslog` (default) and `ISO8601`.

For `Syslog`, the format is `Mmm dd hh:mm:ss`, using the timezone of the syslog device. For example: `Jan 18 16:56:36`.

For `ISO8601`, the basic format is `YYYY-MM-DD`. For example: `2023-01-18`. For extended formats, add hours, minutes, seconds, decimal fractions (optional), and timezone. For example: `2023-01-18T16:56:36.996-08:00`.

Throttling: Throttle rate, in bytes per second. Defaults to `0`, meaning no throttling. Multiple-byte units such as KB, MB, GB etc. are also allowed, e.g., `42 MB`. When throttling is engaged, your **Backpressure behavior** selection determines whether Cribl Edge will handle excess data by blocking it, dropping it, or queueing it to disk.

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Load Balancing Settings

Load balancing is available only when the **Protocol** is set to `TCP`.

Enabling the **Load balancing** toggle replaces the static **General Settings** > **Address** and **Port** fields with the following controls:

Exclude current host IPs: This toggle appears when **Load balancing** is set to `Yes`. It determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Destinations

Use the **Destinations** table to specify a known set of receivers on which to load-balance data. To specify more receivers on new rows, click **Add Destination**. Each row provides the following fields:

Address: Hostname of the receiver. Optionally, you can paste in a comma-separated list, in `<host>:<port>` format.

Port: Port number to send data to on this host.

TLS: Whether to inherit TLS configs from group setting, or disable TLS. Defaults to `inherit`.

TLS servername: Servername to use if establishing a TLS connection. If not specified, defaults to connection host (if not an IP). Otherwise, uses the global TLS settings.

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

Persistent Queue Settings

This section is displayed only for **TCP**, and only when **Backpressure behavior** is set to **Persistent Queue**.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5` GB. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

TLS Settings (Client Side)

Enabled defaults to `No`. When toggled to `Yes`:

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Timeout Settings

These timeout settings apply only to the TCP protocol.

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish, before retrying. Defaults to `10000`.

Write timeout: Amount of time (milliseconds) to wait for a write to complete, before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

The first three options are always displayed:

Octet count framing: When enabled, Cribl Edge prefixes messages with their byte count, regardless of whether the messages are constructed from `message`, `__syslogout`, or `_raw`. When disabled, Cribl Edge omits prefixes, and instead appends a `\n` to messages. RFCs 5425 and 6587 describe how octet counting works in syslog.

Log failed requests to disk: Toggling to `Yes` makes the payloads of any (future) failed requests available for inspection. See Inspecting Payloads to Troubleshoot Closed Connections below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

The following options are added if you enable the General Settings tab's **Load balancing** option:

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that If multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Edge load balancing, IP settings take priority over those from hostnames.)

Max connections: Constrains the number of concurrent receiver connections, per Worker Process, to limit memory utilization. If set to a number > `0`, then on every DNS resolution period, Cribl Edge will randomly select this subset of discovered IPs to connect to. Cribl Edge will rotate IPs in future resolution periods – monitoring weight and historical data, to ensure fair load balancing of events among IPs.

Inspecting Payloads to Troubleshoot Closed Connections

When a downstream receiver closes connections from this Destination (or just stops responding), inspecting the payloads of the resulting failed requests can help you find the cause. For example:

- Suppose you send an event whose size is larger than the downstream receiver can handle.
- Suppose you send an event that has a `number` field, but the value exceeds the highest number that the downstream receiver can handle.

When **Log failed requests to disk** is enabled, you can inspect the payloads of failed requests. Here is how:

1. In the Destination UI, navigate to the **Logs** tab.
2. Find a log entry with a `connection error` message.

3. Expand the log entry.
4. If the message includes the phrase `See payload file for more info`, note the path in the `file` field on the next line.

Now you have the path to the directory where Cribl Edge is storing payloads from failed requests. At the command line, navigate to that directory and inspect any payloads that you think might be relevant.

Understanding Syslog Format Options

Unlike other Cribl Edge Destinations, Syslog applies an additional post-processing step after the Pipeline(s) transform events. This additional step structures the data for compliance with the syslog transport protocol (RFC 3164 and/or RFC 5424) before it is transmitted to downstream services.

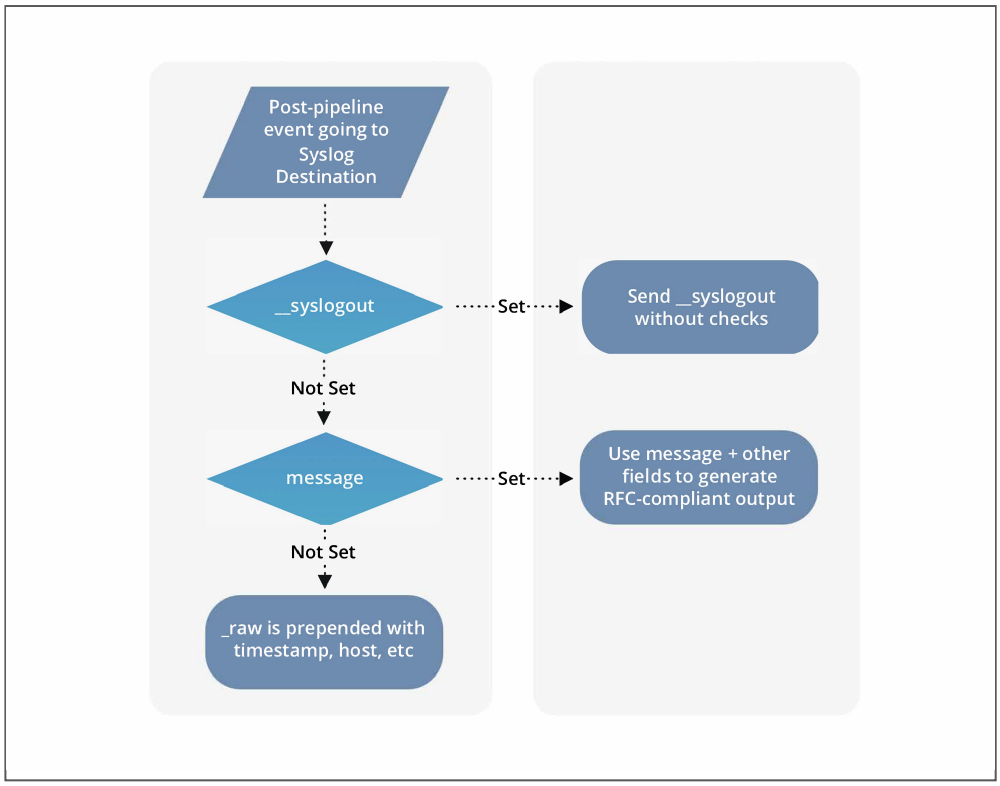
The Syslog Destination's **General Settings** page offers several settings to format the timestamps, to format the message delivering the event, and to set the syslog-specific default settings for Facility, Severity, and App Name.

Below are two examples of RFC-compliant syslog events:

- `<13>Jul 11 10:34:35 testbox testing[42]`
- `<214>1 2022-07-11T18:58:45.000Z testbox testing[42]: foo=bar this=that base=ball gizmo=sprocket`

Cribl Edge offers three different output approaches in the Syslog Destination. The flowchart below summarizes how Cribl Edge determines which approach to use:

- `message`: For ease of use, Cribl recommends using this option. When you define `message`, Cribl Edge discards `_raw`, and composes a new payload using the other syslog-related fields. Cribl Edge automatically processes the values of the `message`, `_time`, `host`, and other fields, and creates an ISO-compliant timestamp and a properly formatted event. To use this method, you must configure `message` and must **not** set `__syslogout`.
- `__syslogout`: When you define `__syslogout`, Cribl Edge sends it as the entire syslog message, discards `_raw`, and ignores all the other fields.
- `_raw` (with or without a header): If you didn't define `message` or `__syslogout`, then Cribl Edge uses the existing `_raw` field as the `message` field, and prepends all the other syslog-related fields to the event.



Syslog output flow

The subsections below walk you through some considerations for each of these options, before you configure this Destination. First, let's take a look at internal fields, since they play a critical role in formatting.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to downstream services. Fields for this Destination:

- `__priority`
- `__facility`
- `__severity`
- `__procid`
- `__appname`
- `__msgid`
- `__host`
- `__syslogout`

Defining `message`

This approach is easiest and least error-prone, because Cribl Edge creates the payload for you. To define the `message`, all you need to supply are the following required fields:

- `__facility`
- `__severity`
- `__host`

Or, to make this even simpler, you can substitute `__priority` for `__facility` and `__severity`. Either way, Cribl Edge will create a new payload using a combination of these required fields. For details on the fields assembled here, see Important Fields below.

Exporting `__syslogout`

As a second approach, if you choose to send `__syslogout` to downstream services, it is exclusive – it becomes the entire syslog message sent. Neither `_raw`, nor any other metadata, will be sent downstream. Also, Cribl Edge will not check to ensure that the value of `__syslogout` is RFC-compliant.

The result will be a proper syslog message only if you hand-build the event yourself. You must manually construct the `__syslogout` payload, starting with `_time`, for all the fields that Cribl Edge would automatically handle with the above `message` option. In particular:

When defining `__syslogout`, you must follow the syslog protocol's RFCs. Otherwise, downstream services will misinterpret or completely ignore the message. Some syslog receivers might drop non-compliant events entirely, or try to "fix" the format by supplying missing fields.

If you have **Octet count framing** enabled for this Destination, Cribl Edge will prepend the number of bytes of the constructed `__syslogout` field to the message before sending it.

The most common uses for the `__syslogout` method are:

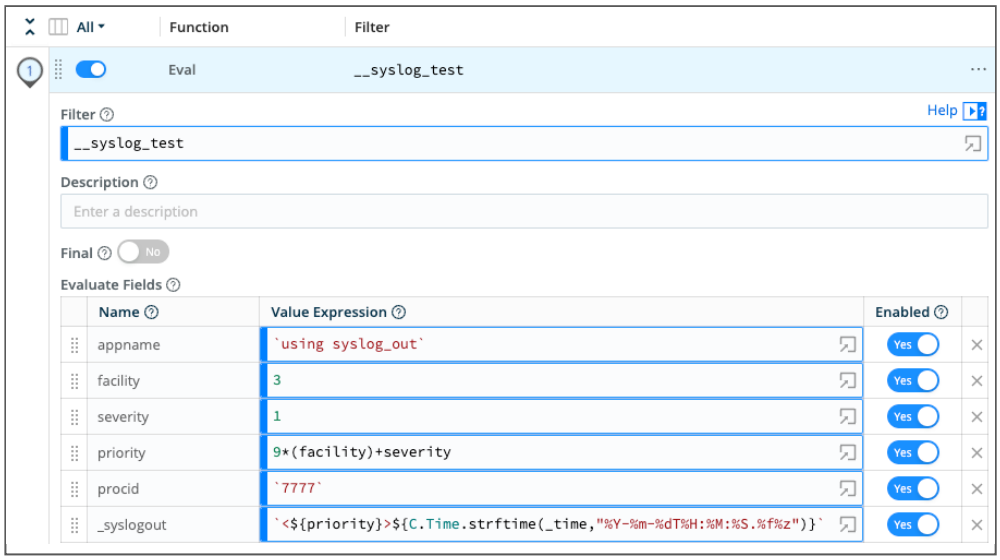
- When the value of `_raw` is already in syslog format as it comes in, and minimal processing is necessary.
- When sending raw TCP data to a destination that doesn't enforce syslog RFCs, such as a raw TCP listener. Cribl Edge currently does not provide a native "raw TCP" Destination. However, in some environments, configuring a Syslog Destination with the TCP protocol and `__syslogout` is an effective method for delivering raw data over TCP.

Constructing `__syslogout`

You'll need to add `_time` to the payload. For example, in an Eval Function, you could use **Evaluate fields** to build up `__syslogout` in the expression below. Here, the `priority` encodes the `severity` + `facility`, according to the syslog protocol.

Name	Value Expression
`priority`	`(8*facility)+severity`
`__syslogout`	`` `<${priority}>${C.Time.strftime(_time,"%Y-%m-%dT%H:%M:%S.%f%z")} ${host} ${appname}[${procid}]: ${_raw}` ``

Here's that example expression in a full Function:

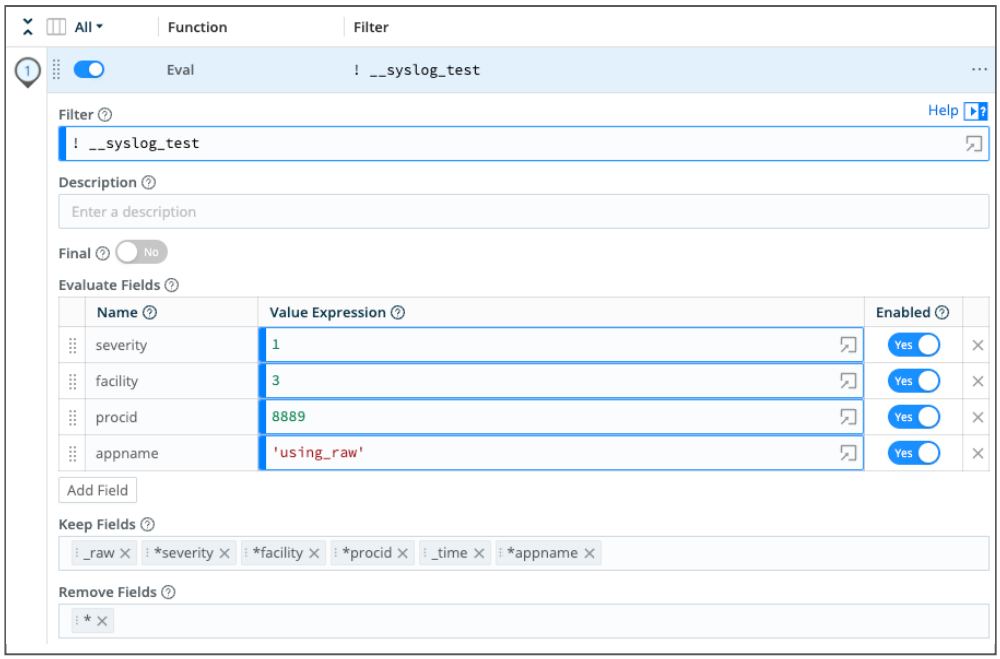


Adding `__syslogout`, `_time` to construct a valid syslog message

Enhancing `_raw` with syslog

A third approach is to add the message content in `_raw`, and construct the syslog "envelope" around `_raw` by including the `severity`, `priority`, `facility`, `procid`, `msgid`, and `appname` fields, as required.

Here's an alternative Eval Function that illustrates this:



Adding syslog decorations to `_raw`

Both Eval Functions are provided in this example Pipeline:

syslog_loop.json

```
{
  "id": "syslog_loop",
  "conf": {
    "output": "default",
    "groups": {},
    "asyncFuncTimeout": 1000,
    "functions": [
      {
        "filter": "true",
        "conf": {
          "mode": "reserialize",
          "type": "json",
          "srcField": "_raw",
          "dstField": "_raw",
          "keep": [
            "resource",
            "path",
            "httpMethod"
          ],
          "remove": []
        },
        "id": "serde",
        "disabled": false
      },
      {
        "filter": "true",
        "conf": {
          "clones": [
            {
              "__syslog_test": "true"
            }
          ]
        },
        "id": "clone",
        "disabled": false
      },
      {
        "filter": "__syslog_test",
        "conf": {
          "add": [
            {
              "name": "appname",
```

```
          "value": "'using_syslogout'"
        },
        {
          "name": "severity",
          "value": "1"
        },
        {
          "name": "facility",
          "value": "3"
        },
        {
          "name": "pri",
          "value": "(8 * facility) + severity"
        },
        {
          "name": "procid",
          "value": "'7777'"
        },
        {
          "name": "__syslogout",
          "value": "`<${pri}>${C.Time.strftime(_time,\"%b %d %H:%M:%S\")} ${hos
        }
      ],
      "keep": [],
      "remove": []
    },
    "id": "eval",
    "disabled": false
  },
  {
    "filter": "! __syslog_test",
    "conf": {
      "add": [
        {
          "name": "severity",
          "value": "1"
        },
        {
          "name": "facility",
          "value": "3"
        },
        {
          "name": "procid",
```

```
            "value": "8889"
          },
          {
            "name": "appname",
            "value": "'using_raw'"
          }
        ],
        "keep": [
          "_raw",
          "*severity",
          "*facility",
          "*procid",
          "_time",
          "*appname"
        ],
        "remove": [
          "*"
        ]
      },
      "id": "eval",
      "disabled": false
    }
  ]
 }
}
```

In this method, Cribl Edge takes the value of `_raw` verbatim, and then prepends a human-readable timestamp, host, and other information.

If you are using `_raw` as the `message` field, make sure you explicitly set the `priority` and `facility` whenever possible. Otherwise, Cribl Edge will use the default values. The acceptable values are defined in the RFCs.

When using `_raw`, you might end up with duplicate fields in the event. For example, even if your event already contains a `timestamp`, Cribl Edge might prepend another `timestamp` to the beginning of `_raw`, without checking whether the data is already present. This can increase event sizes with redundant data. If you use this method, make sure you first strip the prepended (duplicate) fields.

Also, when using `_raw`, this Destination does not check whether there's a valid header defined in the event – it always adds one. Also, because this Destination reformats the data, you might not see the header information when you preview it in Live Capture.

For details on the fields assembled here, see Important Fields below.

Defined `message` and `_raw` Fields

If Cribl Edge's Destination stage receives an event that contains both `message` and `_raw` fields, it will build the syslog package using the `message` field, discarding `_raw`.

The `message` (or `_raw`) field must be an ASCII string in order to be put on the syslog wire. This Destination does not handle JSON objects. Avoid mismatching these types, or else no data might be sent out.

If your intermediate processing – such as a `JSON.parse(_raw)` transformation – has converted the `_raw` field's contents into a JSON object literal, you will need to stringify the result, using a method like `JSON.stringify(_raw)`. You can do so in a post-processing Pipeline attached to the Syslog Destination.

Important Fields

The `Message` and `_raw` prepend methods use additional fields to create the final payload. When using `__syslogout`, Cribl Edge ignores these fields.

The fields below appear in the order generated by a syslog-formatted event.

Sample syslog-formatted event: `<38>1 2022-07-11T22:04:46.000Z testbox app01 [4321] AC-123 - foo=bar this=that base=ball gizmo=sprocket`

- `__facility` or **facility**: Cribl Edge uses this field to calculate priority. The RFC protocol dictates Facility levels. For details, see Facility.
- `__severity` or **severity**: Cribl Edge also uses this field to calculate priority. The RFC protocol dictates Severity levels. For details, see Severity.
- `__priority`: If you configure this field, Cribl Edge will use it and override the `severity` and `facility` values. The priority displays at the beginning of a syslog event, `<38>` in the example above. If you don't configure this field, then Cribl Edge calculates it using the formula: `priority = (8*facility + severity)`.
  For example, if the `facility` is `13` (Security) and the `severity` is `2` (Critical), the `priority` will be `13*8 + 2 = 106`. The `priority` of `<38>` in the example above is `8*4(Facility of auth) + 6 (Severity of info)`.
- `_time`: The value of `_time` in Cribl events is in epoch format, but the syslog RFCs dictate that each event's timestamp is must be in human-readable format. When defining a Syslog Destination, you can have an option to use the ISO8601 (recommended) or the syslog format. ISO8601 defines the method for specifying time zone and year, whereas the older syslog format lacks this information. The example above shows the ISO8601 format, listed after the `priority` (`<13>`) and `version`.

- `__host` OR **host**: the value for the required host field in a syslog event, following the timestamp. `testbox` in the example above.
- `__appname` or **appname**: The required application name. `app01` in the example above. This is typically the name of the daemon or process that is logging any given event.
- `__procid` or **procid**: This optional field is the Process ID. `4321` in the example above. Use a numeric value for this field, optionally this may be surrounded with brackets . Cribl Edge will automatically adjust the spaces and syntax to ensure RFC-compliant formatting.
- `__msgid` or **msgid**: This optional field is the Message ID. `AC-123` in the example above.

For each pair of attribute names (above), Cribl Edge uses the values as specified here:

1. The event contains both `__<key>` and `<key>`: Cribl Edge uses the value of `__<key>` and ignores `<key>` if also present. For example, if the event contains `__host`, the value of **host** will be ignored if also present.
2. The event contains `<key>`: Cribl Edge uses the value of `<key>`. For example, if **host** is set, the value is applied.
3. The event contains neither `__<key>` nor `<key>`, Cribl Edge uses the default values configured in the Syslog Destination.

11.25. TCP JSON

Cribl Edge supports sending data over TCP in JSON format.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output Data in TCP JSON Format

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **TCP JSON**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **TCP JSON**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Destination definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Load balancing: When enabled (default), lets you specify multiple destinations. See Load Balancing Settings below. The following two fields appear **only** with the `No` setting.

Address: Hostname of the receiver.

Port: Port number to connect to on the host.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: In the resulting **Auth token** field, you can optionally enter an auth token to use in the connection header.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the `authToken` header field value described above. A **Create** link is

available to store a new, reusable secret.

Optional Settings

Exclude current host IPs: This toggle appears when **Load balancing** is set to `Yes`. It determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Compression: Codec to use to compress the data before sending. Defaults to `Gzip`.

Throttling: Throttle rate, in bytes per second. Defaults to `0`, meaning no throttling. Multiple-byte units such as KB, MB, GB etc. are also allowed, e.g., `42 MB`. When throttling is engaged, your **Backpressure behavior** selection determines whether Cribl Edge will handle excess data by blocking it, dropping it, or queueing it to disk.

Backpressure behavior: Specifies whether to block, drop, or queue events when all receivers are exerting backpressure. Defaults to `Block`. See Persistent Queue Settings below.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Load Balancing Settings

Enabling the **Load balancing** toggle replaces the static **General Settings** > **Address** and **Port** fields with the following controls:

Destinations

Use the **Destinations** table to specify a known set of receivers on which to load-balance data. To specify more receivers on new rows, click **Add Destination**. Each row provides the following fields:

Address: Hostname of the receiver. Optionally, you can paste in a comma-separated list, in `<host>:<port>` format.

Port: Port number to send data to on this host.

TLS: Whether to inherit TLS configs from group setting, or disable TLS. Defaults to `inherit`.

TLS servername: Servername to use if establishing a TLS connection. If not specified, defaults to connection host (if not an IP). Otherwise, uses the global TLS settings.

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

TLS Settings (Client Side)

Use TLS defaults to `No`. When toggled to `Yes`:

Autofill?: This setting is experimental.

Validate server certs: Reject certificates that are not authorized by a CA in the **CA certificate path**, or by another trusted CA (e.g., the system's CA). Defaults to `Yes`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Timeout Settings

Connection timeout: Amount of time (in milliseconds) to wait for the connection to establish before retrying. Defaults to `10000`.

Write timeout: Amount of time (in milliseconds) to wait for a write to complete before assuming connection is dead. Defaults to `60000`.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Advanced Settings

Log failed requests to disk: Toggling to `Yes` makes the payloads of any (future) failed requests available for inspection. See Inspecting Payloads to Troubleshoot Closed Connections below.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Setting General Settings > **Load balancing** to `Yes` adds the following settings:

DNS resolution period (seconds): Re-resolve any hostnames after each interval of this many seconds, and pick up destinations from A records. Defaults to `600` seconds.

Load balance stats period (seconds): Lookback traffic history period. Defaults to `300` seconds. (Note that If multiple receivers are behind a hostname – i.e., multiple A records – all resolved IPs will inherit the weight of the host, unless each IP is specified separately. In Cribl Edge load balancing, IP settings take priority over those from hostnames.)

Max connections: Constrains the number of concurrent receiver connections, per Worker Process, to limit memory utilization. If set to a number > `0`, then on every DNS resolution period, Cribl Edge will randomly select this subset of discovered IPs to connect to. Cribl Edge will rotate IPs in future resolution periods – monitoring weight and historical data, to ensure fair load balancing of events among IPs.

Inspecting Payloads to Troubleshoot Closed Connections

When a downstream receiver closes connections from this Destination (or just stops responding), inspecting the payloads of the resulting failed requests can help you find the cause. For example:

- Suppose you send an event whose size is larger than the downstream receiver can handle.
- Suppose you send an event that has a `number` field, but the value exceeds the highest number that the downstream receiver can handle.

When **Log failed requests to disk** is enabled, you can inspect the payloads of failed requests. Here is how:

1. In the Destination UI, navigate to the **Logs** tab.
2. Find a log entry with a `connection error` message.
3. Expand the log entry.
4. If the message includes the phrase `See payload file for more info`, note the path in the `file` field on the next line.

Now you have the path to the directory where Cribl Edge is storing payloads from failed requests. At the command line, navigate to that directory and inspect any payloads that you think might be relevant.

Format

TCP JSON events are sent in newline-delimited JSON format, consisting of:

1. A header line. Can be empty, e.g.: `{}`. If **Auth Token** is enabled, the token will be included here as a field called `authToken`. In addition, if events contain common fields, they will be included here under `fields`.
2. A JSON event/record per line.

See an example in our TCP JSON Source topic.

11.26. Wavefront

Cribl Edge supports sending events to Wavefront analytics.

Type: Streaming | TLS Support: Yes | PQ Support: Yes

Configuring Cribl Edge to Output to Wavefront

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Wavefront**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Wavefront**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this Wavefront definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Domain name: WaveFront domain name, e.g., `longboard`. Required.

Authentication Settings

Use the **Authentication method** drop-down to select one of these options:

- **Manual**: Displays an **API key** field for you to enter your Wavefront API auth token. See Wavefront's Generating an API Token topic.
- **Secret**: This option exposes an **Auth token (text secret)** drop-down, in which you can select a stored secret that references the API auth token described above. A **Create** link is available to store a new, reusable secret.

Optional Settings

Backpressure behavior: Select whether to block, drop, or queue events when all receivers are exerting backpressure. (Causes might include a broken or denied connection, or a rate limiter.) Defaults to `Block`.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.

Status Code	Action
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Toggle to `Yes` to reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (e.g., the system's CA).

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. The actual request body size might exceed the specified value because the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Flush period (sec): Maximum time between requests. Low values can cause the payload size to be smaller than the configured **Max body size**. Defaults to `1` second.

Extra HTTP headers: Click **Add Header** to insert extra headers as **Name**/**Value** pairs. Values will be sent encrypted.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Proxying Requests

If you need to proxy HTTP/S requests, see System Proxy Configuration.

Notes About Wavefront

For details on integrating with Wavefront, see these Wavefront resources:

- Direct Data Ingestion, and adjacent topics on Wavefront Proxies.
- Wavefront Data Format.

11.27. Webhook

Cribl Edge can send log and metric events to webhooks and other generic HTTP endpoints.

Type: Streaming | TLS Support: Configurable | PQ Support: Yes

Configuring Cribl Edge to Output via HTTP

From the top nav, click **Manage**, then select a **Fleet** to configure. Next, you have two options:

To configure via the graphical QuickConnect UI, click Routing > QuickConnect (Stream) or Collect (Edge). Next, click **Add Destination** at right. From the resulting drawer's tiles, select **Webhook**. Next, click either **Add Destination** or (if displayed) **Select Existing**. The resulting drawer will provide the options below.

Or, to configure via the Routing UI, click **Data** > **Destinations** (Stream) or **More** > **Destinations** (Edge). From the resulting page's tiles or the **Destinations** left nav, select **Webhook**. Next, click **Add Destination** to open a **New Destination** modal that provides the options below.

General Settings

Output ID: Enter a unique name to identify this HTTP output definition. If you clone this Destination, Cribl Edge will add `-CLONE` to the original **Output ID**.

Load balancing: When enabled (default), lets you specify multiple Webhook URLs and load weights. With the default `No` setting, if you notice that Cribl Edge is not sending data to all possible IP addresses, enable **Advanced Settings** > **Round-robin DNS**.

Webhook URL: Endpoint URL to send events to.

Webhook URLs

Use the **Webhook URLs** table to specify a known set of receivers on which to load-balance data. To specify more receivers on new rows, click **Add URL**. Each row provides the following fields:

Webhook LB URL: Specify the Webhook URL to send events to – for example, `http://webhook:8000/`

Load weight: Set the relative traffic-handling capability for each connection by assigning a weight (> `0`). This column accepts arbitrary values, but for best results, assign weights in the same order of magnitude to all

connections. Cribl Edge will attempt to distribute traffic to the connections according to their relative weights.

The final column provides an `X` button to delete any row from the table.

For details on configuring all these options, see About Load Balancing.

When you first enable load balancing, or if you edit the load weight once your data is load–balanced, give the logic time to settle. The changes might take a few seconds to register.

Optional Settings

Exclude current host IPs: This toggle appears when **Load balancing** is set to `Yes`. It determines whether to exclude all IPs of the current host from the list of any resolved hostnames. Defaults to `No`, which keeps the current host available for load balancing.

Method: The HTTP verb to use when sending events. Defaults to `POST`. Change this to `PUT` or `PATCH` where required by the endpoint.

Format: The format in which to send out events. One of:

- `NDJSON`: Newline-delimited JSON (the default).
- `JSON Array`: Arrays in JSON-parseable format.
- `Custom`: Events and event batches (payloads) whose format you define using JavaScript **expressions**. See Custom Format below.
- `Advanced`: Events and event batches (payloads) whose format you define using JavaScript **code**. Enables the Webhook Destination to integrate with APIs not otherwise supported in Cribl Edge. See Advanced Format below.

Backpressure behavior: Whether to block, drop, or queue events when all receivers are exerting backpressure.

Tags: Optionally, add tags that you can use to filter and group Destinations in Cribl Edge's **Manage Destinations** page. These tags aren't added to processed events. Use a tab or hard return between (arbitrary) tag names.

TLS Settings (Client Side)

Enabled Defaults to `No`. When toggled to `Yes`:

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates in (PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Authentication

Select one of the following options for authentication:

- **None**: Don't use authentication.
- **Auth token**: Use HTTP token authentication. In the resulting **Token** field, enter the bearer token that must be included in the HTTP authorization header.
- **Auth token (text secret)**: This option exposes a **Token (text secret)** drop-down, in which you can select a stored text secret that references the bearer token described above. A **Create** link is available to store a new, reusable secret.
- **Basic**: Displays **Username** and **Password** fields for you to enter HTTP Basic authentication credentials.
- **Basic (credentials secret)**: This option exposes a **Credentials secret** drop-down, in which you can select a stored text secret that references the Basic authentication credentials described above. A **Create** link is available to store a new, reusable secret.
- **OAuth**: Configure authentication via the OAuth protocol. Exposes multiple extra options – see OAuth Authentication below.

OAuth Authentication

Selecting **OAuth** exposes the following additional fields:

- **Login URL**: Endpoint for the OAuth API call, which is expected to be a POST call.

- **OAuth secret**: Secret parameter value to pass in the API call's request body.
- **OAuth secret parameter name**: Secret parameter name to pass in API call's request body.
- **Token attribute name**: Name of the auth token attribute in the OAuth response. Can be top-level (for example, `token`); or nested, using a period (for example, `data.token`).
- **Authorize expression**: JavaScript expression used to compute the Authorization header to pass in requests. Uses `${token}` to reference the token obtained from the login POST request, for example: `` `Bearer ${token}` ``.
- **Refresh interval (secs.)**: How often to refresh the OAuth token. Default `3600` seconds (1 hour); minimum `1` second; maximum `300000` seconds (about 83 hours).
- **OAuth parameters**: Optionally, click **Add parameter** for each additional parameter you want to send in the OAuth login request. Cribl Edge will combine the secret with these parameters, and will send the URL-encoded result in a POST request to the endpoint specified in the **Login URL**. We'll automatically add the `Content-Type` header `application/x-www-form-urlencoded` when sending this request.
  In each row of the resulting table, enter the parameter's **Name**. The corresponding **Value** is a JavaScript expression to compute the parameter's value. This can also evaluate to a constant. Values not formatted as expressions – for example, `id` instead of `` `${id}` `` – will be evaluated as strings.
- **Authentication headers**: Optionally, click **Add Header** for each custom auth header you want to define and send in the OAuth login request. Stream will automatically add the `Content-Type` header `application/x-www-form-urlencoded` when sending this request.
  In each row of the resulting table, enter the custom OAuth header's **Name**. The corresponding **Value** is a JavaScript expression to compute the header's value. This can also evaluate to a constant. Values not formatted as expressions – for example, `id` instead of `` `${id}` `` – will be evaluated as strings.

Persistent Queue Settings

This tab is displayed when the **Backpressure behavior** is set to **Persistent Queue**.

On Cribl-managed Cribl.Cloud Workers (with an Enterprise plan), this tab exposes only the destructive **Clear Persistent Queue** button (described below in this section). A maximum queue size of 1 GB disk space is automatically allocated per PQ-enabled Destination, per Worker Process. The 1 GB limit is on outbound uncompressed data, and no compression is applied to the queue.

This limit is not configurable. If the queue fills up, Cribl Edge will block outbound data. To configure the queue size, compression, queue-full fallback behavior, and other options below, use a hybrid Group.

Max file size: The maximum data volume to store in each queue file before closing it. Enter a numeral with units of KB, MB, etc. Defaults to `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume on each Worker Process. Once this limit is reached, this Destination will stop queueing data and apply the **Queue-full** behavior. Required, and defaults to `5 GB`. Accepts positive numbers with units of `KB`, `MB`, `GB`, etc. Can be set as high as `1 TB`, unless you've configured a different **Max PQ size per Worker Process** in Fleet Settings.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Queue-full behavior: Whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the **Block** option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Processing Settings

Post-Processing

Pipeline: Pipeline to process data before sending the data out using this output.

System fields: A list of fields to automatically add to events that use this output. By default, includes `cribl_pipe` (identifying the Cribl Edge Pipeline that processed the event). Supports wildcards. Other

options include:

- `cribl_host` – Cribl Edge Node that processed the event.
- `cribl_input` – Cribl Edge Source that processed the event.
- `cribl_output` – Cribl Edge Destination that processed the event.
- `cribl_route` – Cribl Edge Route (or QuickConnect) that processed the event.
- `cribl_wp` – Cribl Edge Worker Process that processed the event.

Retries

Honor Retry-After header: Whether to honor a `Retry-After` header, provided that the header specifies a delay no longer than 180 seconds. Cribl Edge limits the delay to 180 seconds even if the `Retry-After` header specifies a longer delay. When enabled, any `Retry-After` header received takes precedence over all other options configured in the **Retries** section. When disabled, all `Retry-After` headers are ignored.

Settings for failed HTTP requests: When you want to automatically retry requests that receive particular HTTP response status codes, use these settings to list those response codes.

For any HTTP response status codes that are not explicitly configured for retries, Cribl Edge applies the following rules:

Status Code	Action
Greater than or equal to `400` and less than or equal to `500`.	Drop the request.
Greater than `500`.	Retry the request.

Upon receiving a response code that's on the list, Cribl Edge first waits for a set time interval called the **Pre-backoff interval** and then begins retrying the request. Time between retries increases based on an exponential backoff algorithm whose base is the **Backoff multiplier**, until the backoff multiplier reaches the **Backoff limit (ms)**. At that point, Cribl Edge continues retrying the request without increasing the time between retries any further.

By default, this Destination has no response codes configured for automatic retries. For each response code you want to add to the list, click **Add Setting** and configure the following settings:

- **HTTP status code**: A response code that indicates a failed request, for example `429 (Too Many Requests)` or `503 (Service Unavailable)`.
- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).

- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Retry timed-out HTTP requests: When you want to automatically retry requests that have timed out, toggle this control on to display the following settings for configuring retry behavior:

- **Pre-backoff interval (ms)**: The amount of time to wait before beginning retries, in milliseconds. Defaults to `1000` (one second).
- **Backoff multiplier**: The base for the exponential backoff algorithm. A value of `2` (the default) means that Cribl Edge will retry after 2 seconds, then 4 seconds, then 8 seconds, and so on.
- **Backoff limit (ms)**: The maximum backoff interval Cribl Edge should apply for its final retry, in milliseconds. Default (and minimum) is `10,000` (10 seconds); maximum is `180,000` (180 seconds, or 3 minutes).

Advanced Settings

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (for example, the system's CA). Defaults to `Yes`.

Round-robin DNS: Toggle to `Yes` to use round-robin DNS lookup across multiple IPv6 addresses. When a DNS server returns multiple addresses, this will cause Cribl Edge to cycle through them in the order returned.

Compress: Compresses the payload body before sending. Defaults to `Yes` (recommended).

Keep alive: By default, Cribl Edge sends `Keep-Alive` headers to the remote server and preserves the connection from the client side up to a maximum of 120 seconds. Toggle this off if you want Cribl Edge to close the connection immediately after sending a request.

Request timeout: Amount of time (in seconds) to wait for a request to complete before aborting it. Defaults to `30`.

Request concurrency: Maximum number of concurrent requests per Worker Process. When Cribl Edge hits this limit, it begins throttling traffic to the downstream service. Defaults to `5`. Minimum: `1`. Maximum: `32`.

Max body size (KB): Maximum size of the request body before compression. Defaults to `4096` KB. You can set this limit to as high as 500 MB (`512000` KB). Be aware that high values can cause high memory usage per Worker Node, especially if a dynamically constructed URL (see Internal Fields) causes this Destination to send events to more than one URL. The actual request body size might exceed the specified value because

the Destination adds bytes when it writes to the downstream receiver. Cribl recommends that you experiment with the **Max body size** value until downstream receivers reliably accept all events.

Max events per request: Maximum number of events to include in the request body. The `0` default allows unlimited events.

Flush period (sec): Maximum time between requests. Low values could cause the payload size to be smaller than its configured maximum. Defaults to `1`.

Extra HTTP headers: Name-value pairs to pass to all events as additional HTTP headers. Values will be sent encrypted. You can also add headers dynamically on a per-event basis in the `__headers` field. See Internal Fields below.

Failed request logging mode: Use this drop-down to determine which data should be logged when a request fails. Select among `None` (the default), `Payload`, or `Payload + Headers`. With this last option, Cribl Edge will redact all headers, except non-sensitive headers that you declare below in **Safe headers**.

Safe headers: Add headers to declare them as safe to log in plaintext. (Sensitive headers such as `authorization` will always be redacted, even if listed here.) Use a tab or hard return to separate header names.

Environment: If you're using GitOps, optionally use this field to specify a single Git branch on which to enable this configuration. If empty, the config will be enabled everywhere.

Custom Format

Choosing `Custom` from the **General Settings** > **Format** drop-down exposes the additional fields described in this section. **Source expression** and most of the rest of the controls define the format of individual events – that is, what data you want to make available from the event. **Batch expression** defines the wrapper object for batched events – that is, how to package events within the request payload. Both expressions **must** be enclosed in backticks.

- **Source expression**: JavaScript expression to evaluate on every event; Cribl Edge will send the result of that evaluation instead of the original event. Sample expression: `` `${fieldA}, ${fieldB}` `` (with literal backticks). Defaults to `__httpOut` – that is, the value of the `__httpOut` field. Use the button at right to open a validation modal.
- **Drop when null**: If toggled to `Yes`, Cribl Edge will drop events when the **Source expression** evaluates to `null`.
- **Event delimiter**: Delimiter string to insert between events. Defaults to the newline character (`\n`). Cannot be a space (this will be converted to `\n`).
- **Content type**: Content type to use for requests. Defaults to `application/x-ndjson`. Any content types set in **Advanced Settings** > **Extra HTTP headers** will override this entry.

- **Batch expression**: Expression specifying how to format the payload for each batch. Defines a wrapper object in which to include the formatted events, such as a JSON document. This enables requests to APIs that require such objects. To reference the events to send, use the `${events}` variable. An example expression to send the batch inside a JSON object would be: `{"items" : [${events}] }`.

Custom Format Example

Suppose that the API you're sending events to requires request payloads in the following form:

```
{
   "Events":[
      {
         "one":"1",
         "two":"2"
      },
      {
         "one":"1.1",
         "two":"2.2"
      }
   ]
}
```

And here's how the events coming into your Webhook Destination look – note that the field names need to change, and `field3` and `field4` need to be dropped:

```
{ "field1": 1, "field2": 2, "field3": 3, "field4": 4 }
{ "field1": 1.1, "field2": 2.2, "field3": 3.3, "field4": 4.4 }
```

To send these events in the required form, you'd do the following:

- Set the **Source expression** to `` `{ "one": "${field1}", "two": "${field2}" }` `` – note that this renames and sends only the desired fields (`one` and `two`).
- Enter the comma (`,`) as your **Event delimiter**.
- Set the **Batch expression** to `` `{ "Events": [${events}] }` ``.

Advanced Format

Choosing `Advanced` from the **General Settings** > **Format** drop-down exposes the JavaScript code blocks described in this section. You can define a format using either one of the code blocks, or both.

- **Format inbound event**: JavaScript code block that receives incoming events (as JavaScript objects) and converts them into strings. Multiple strings are concatenated to form a batch, the batches constituting the payloads of the HTTP requests that the Destination sends to the downstream API.
- **Format outbound payload**: JavaScript code block that receives batches before they are sent out as HTTP request payloads. This gives you the opportunity to modify the payload format from the way it arrives – as a string of concatenated events – to whatever format the downstream API requires.

Code Block Restrictions

Generally speaking, anything forbidden in JavaScript strict mode is forbidden in the code blocks. Specifically, the following are **not allowed**:

- `console`, `eval`, `uneval`, `Function (constructor)`, `Promises`, `setTimeout`, `setInterval`, `global`, `globalThis`, `window`, and `set`.

Code blocks **can** include `for` loops, `while` loops, and JavaScript methods such as `map`, `reduce`, `forEach`, `some`, and `every`. For further details, see Supported JavaScript Options.

Only skilled JavaScript developers should use the Advanced format. This is to avoid unintended results – such as creating infinite loops, or otherwise failing to return – that could needlessly add to your throughput burden.

Advanced Format Syntax

Here are the syntax conventions for manipulating events and payloads within the code blocks.

Format inbound event Syntax

- `__e`: References the event object.
- `__e['fieldName']`: References an individual field in the event object.
- `__e.asJSON(): string`: Method to safely convert the object to JSON by removing internal content that is not relevant for a downstream system. **Always** use this method instead of directly converting the object to JSON string using `JSON.stringify(__e)`. This avoids circular references that can cause `JSON.stringify(__e)` to fail.

Format outbound event Syntax

In addition to the syntax conventions defined above, the **Format outbound event** code block has its own special conventions:

- `__e['__payload']`: Accesses the `payload` attribute of the JavaScript object that the code block receives. The `__payload` attribute contains the actual batch (payload) to be formatted.

- `__e['__payloadOut']`: After you modify the batch (payload) as desired, assign it to this variable. If the `__payloadOut` attribute is not present, the code block sends the payload as-is. See the JSON Array Example below.

Advanced Format Examples

The **Format inbound event** and **Format outbound payload** code blocks are populated by placeholder JavaScript code. This code expresses the basic flow you must understand to define an Advanced format. Let's examine the code – then you'll be able to see how defining any given Advanced format is just a matter of manipulating events and payloads within the basic flow.

The Format inbound event Code Block

The event to format appears in the first line, referenced as `__e: CriblEvent`. First, you serialize the event as a string; the simplest way to do that is with the method `e.asJSON()`. Then, you assign it back to the object in a variable named `__eventOut`.

```
formatEvent (__e : CriblEvent) {
  // TODO: Serialize the event to a string here.
  let formattedEvent = __e.asJSON();

  // Assign formatted event to __eventOut.
  __e['__eventOut'] = formattedEvent;
}
```

At this point, the `formatEvent()` function ensures that each incoming event will be serialized as a string. (In the more elaborate examples below, the format of that string will be manipulated.) Each new event, in string form, is appended to the last, building a longer string that will become a **batch**. How many events can be in a batch is governed by **Advanced Settings** > **Max events per request**.

Once a batch is complete, the next code block can operate on it.

The Format outbound payload Code Block

Now the `formatPayload ()` function receives an event whose `payload` attribute is the batch produced by the previous code block. You assign that attribute to the `formattedPayload` variable. If you want to manipulate the batch, do that here. Once you have transformed the batch, you assign it to the `__payloadOut` variable, and the Webhook Destination sends the batch out as the payload of an HTTP request.

Transforming batches is not required. To send batches out exactly how they leave the **Format inbound event** code block, just leave the **Format outbound payload** code block blank.

```
formatPayload: (__e : CriblEvent) {
  // Access payload to format here
  let formattedPayload = __e['payload']

  // TODO: Format payload here

  // Assign formatted payload to payloadOut
  __e['payloadOut'] = formattedPayload
}
```

JSON Array Example

Suppose you want to communicate with an API that accepts data in JSON array format.

Use the **Format inbound event** code block to format the events:

```
__e['__eventOut'] = `${__e.asJSON()},` // Serialize to JSON and add trailing comma
```

Use the **Format outbound event** code block to format the payload:

```
// Remove trailing comma and wrap payload in array
__e['__payloadOut'] = `[${payload.substring(0,payload.length-1)}]`
```

Elasticsearch Bulk API Example

The Elasticsearch Bulk API requires each event to be formatted with two lines of output. While this cannot be done with the Webhook Destination's other format options, it is doable with Advanced Format. (The Elasticsearch API Source, Elasticsearch Destination and Elastic Cloud Destination **do** work with the Elasticsearch Bulk API.)

Here's how Bulk API events must be formatted, structurally:

```
{ <action>:  { <metadata> } }
{ <serializedEvent> }
```

Here's an example of an event formatted for Bulk API:

```
{ "index" : { "_index" : "test", "_id" : "1" } }
{ "field1" : "value1", "field2" "value2", "field3" : "value3" }
```

Use the **Format inbound event** code block to format the events:

```
let time=__e['_time'] || Date.now();
__e['_time'] = undefined;
__e['@timestamp'] = C.Time.strftime(time, '%Y-%m-%dT%H:%M:%S.%LZ');
if(typeof __e['host'] === 'string') {
  const host = __e['__host'] || {};
  host.name = __e['host'];
  __e['host'] = host;
}
__e['__eventOut'] = `{"create":{"_index":"${__e['_index'] || 'cribl'}", "pipeline"
```

Leave the **Format outbound event** code block blank.

Splunk HEC Example

Use the **Format inbound event** code block to format the events:

```
    const eventIn = __e;
    const time = eventIn._time;
    const event = eventIn._raw;
    const {host, source, sourcetype, index} = eventIn;
    eventIn._time = undefined;
    eventIn._raw = undefined;
    eventIn.host = undefined;
    eventIn.source = undefined;
    eventIn.sourcetype = undefined;
    eventIn.index = undefined;
    const fields = {};
    const obj2send = {time, index, host, source, sourcetype, event, fields};
    for (const key of ['index', 'host', 'source', 'sourcetype']) {
      // only convert if the key/val is present and is not a string
      const oldValue = obj2send[key];
      if (oldValue != null && typeof oldValue !== 'string') {
        obj2send[key] = JSON.stringify(oldValue);
      }
    }
    for (let key of Object.keys(eventIn)) {
      if(key.startsWith('__'))
        continue;
      let val = eventIn[key];
      if(val === undefined)
        continue;
      if(typeof val === 'object')
        val = JSON.stringify(val);
      obj2send.fields[key] = val;
    }
  __e['__eventOut'] = `${JSON.stringify(obj2send)}\n`;
```

Leave the **Format outbound event** code block blank.

Internal Fields

Cribl Edge uses a set of internal fields to assist in forwarding data to a Destination.

Fields for this Destination:

- `__criblMetrics`

- `__url`
- `__headers`

If an event contains the internal field `__criblMetrics`, Cribl Edge will send it to the HTTP endpoint as a metric event. Otherwise, Cribl Edge will send it as a log event.

When **Load balancing** is disabled, if an event that contains the internal field `__url`, that field's value will override the **General Settings** > **Webhook URL** value. This way, you can determine the endpoint URL dynamically.

For example, you could create a Pipeline containing an Eval Function that adds the `__url` field, and connect that Pipeline to your Webhook Destination. Configure the Eval Function to set `__url`'s value to a URL that varies depending on a global variable, or on some property of the event, or on some other dynamically-generated value that meets your needs.

If an event contains the internal field `__headers`, that field's value will be a JSON object containing Name-value pairs, each of which defines a header. By creating Pipelines that set the value of `__headers` according to conditions that you specify, you can add headers dynamically.

For example, you could create a Pipeline containing Eval Functions that add the `__headers` field, and connect that Pipeline to your Webhook Destination. Configure the Eval Functions to set `__headers` values to Name-value pairs that vary depending on some properties of the event, or on dynamically-generated values that meet your needs.

Here's an overview of how to add headers:

- To add "dynamic" (a.k.a. "custom") headers to some events but not others, use the `__headers` field.
- To define headers to add to **all** events, use **Advanced Settings** > **Extra HTTP Headers**.
- An event can include headers added by both methods. So if you use `__headers` to add `{ "api-key": "foo" }` and **Extra HTTP Headers** to add `{ "goat": "Kid A" }`, you'll get both headers.
- Headers added via the `__headers` field take precedence. So if you use `__headers` to add `{ "api-key": "foo" }` and **Extra HTTP Headers** to add `{ "api-key": "bar" }`, you'll get only one header, and that will be `{ "api-key": "foo" }`.

Use Cases

See these examples of configuring a Webhook Destination to integrate with specific services:

- Azure Analytics/Sentinel Integration
- Webhook/BigPanda Integration

- Webhook/Slack Integration

Notes on HTTP-Based Outputs

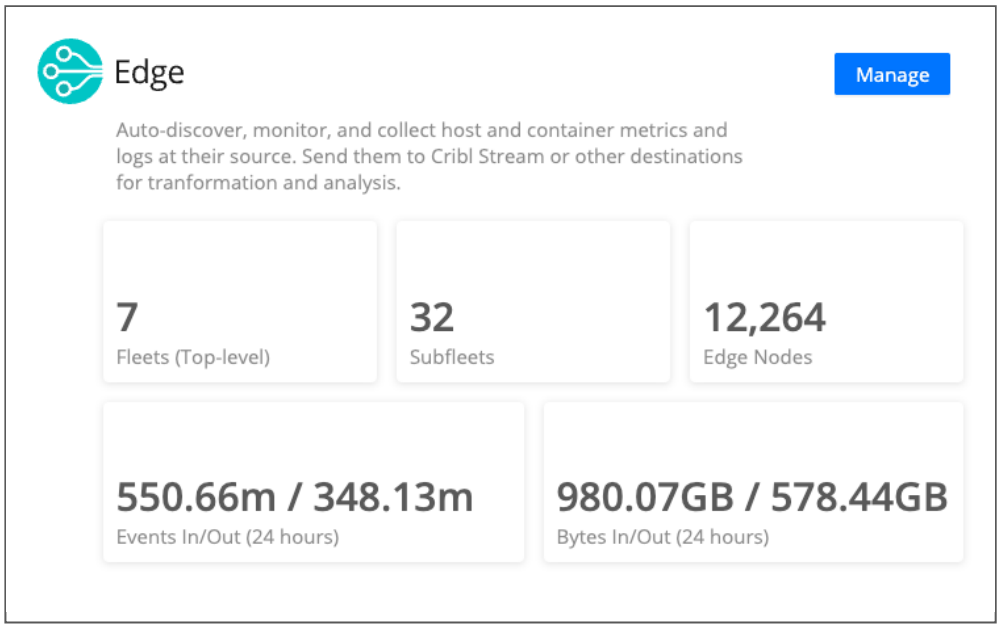
- To proxy outbound HTTP/S requests, see System Proxy Configuration.
- Cribl Edge will attempt to use keepalives to reuse a connection for multiple requests. After two minutes of the first use, it will throw away the connection and attempt a new one. This is to prevent sticking to a particular destination when there is a constant flow of events.
- If the server does not support keepalives (or if the server closes a pooled connection while idle), Cribl Edge will establish a new connection for the next request.
- When resolving the Destination's hostname, Cribl Edge will pick the first IP in the list for use in the next connection. Enable **Round-robin DNS** to better balance distribution of events among destination cluster nodes.

12. Monitoring Health And Metrics

To get an operational view of your Cribl Edge deployment, consult the following resources.

Landing Pages

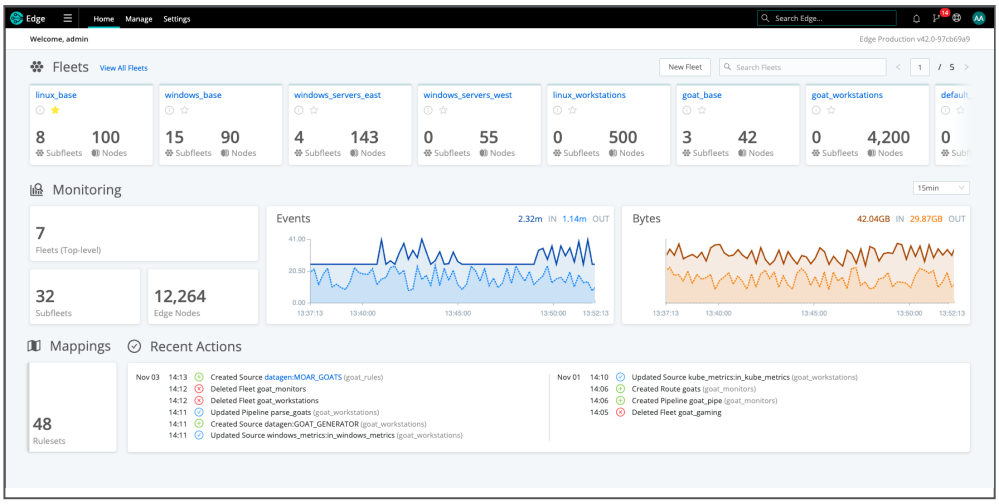
When you first log into your distributed deployment, you will be greeted with tiles prompting you to choose a product. The Cribl Edge tile displays basic configuration details, including number of Fleets, Subfleets, Edge Nodes, and events and bytes over time.



Manage Edge Tile

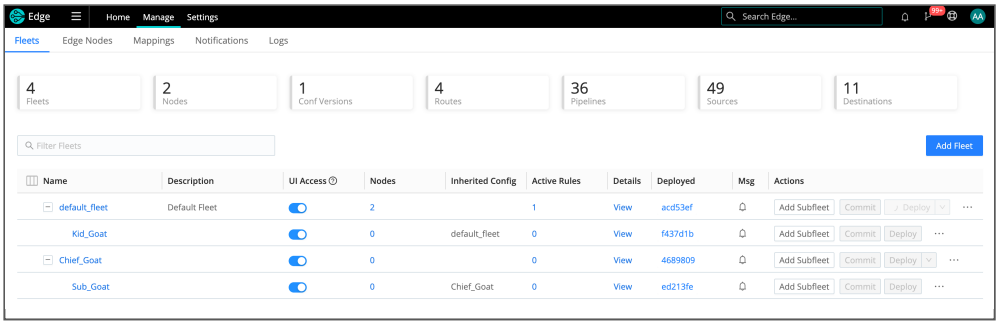
Select **Manage** to navigate to the Cribl Edge **Home** page, which shows aggregate data for all Fleets, Subfleets and Edge Nodes. The charts display information about traffic in and out of the system.

On top of the **Bytes** chart, you can change the display's granularity from the default `last 5 min`, selecting a variety of time ranges from `1 min` up to `1 day` (covering the preceding 24 hours).



Cribl Edge Home page

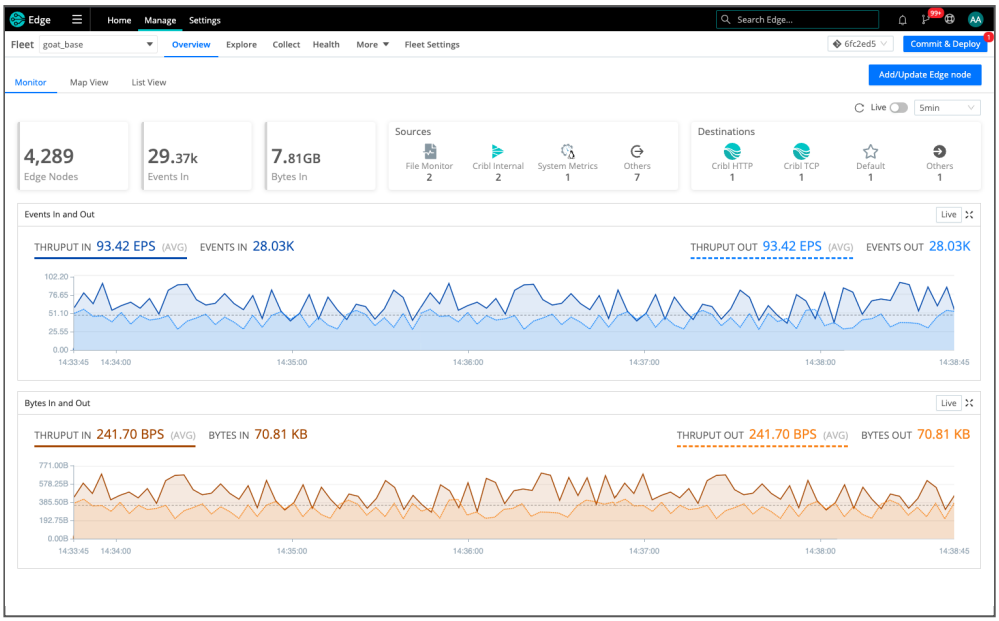
Select **Manage** from the top nav to view the **Fleets Landing** page. Here you can access the tabs for more information about your **Fleets** (and **Subfleets**), **Edge Nodes**, **Mappings**, Notifications, and **Logs**.



Manage Fleets

The **Manage Fleets** page gives you access to more information about your **Fleets** (and **Subfleets**), **Edge Nodes**, **Mappings**, Notifications, and **Logs**.

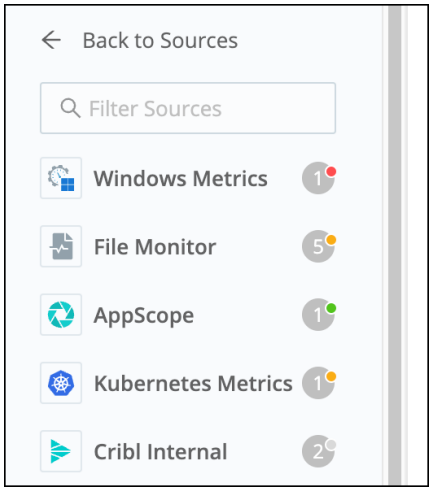
You can click a **Fleet** link to isolate individual Fleets, or use the **Search** bar to locate your Fleet.



Fleet Landing page

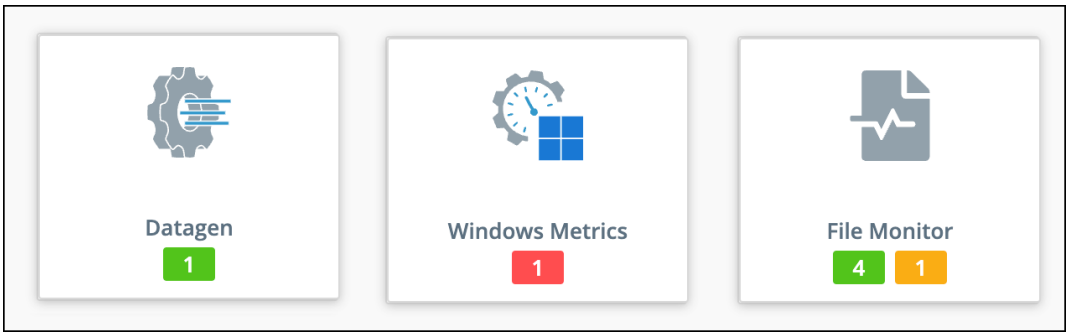
Source and Destination Health Status

When you add or manage configured Sources and Destinations, you'll see a colored health status (if enabled). In the list view, this status is a green, orange, red, or gray dot next to the name, along with a gray circle indicating the number of instances you have configured.



Source list view

In the tile view, colors are applied to the squares that indicate the number of instances you have configured.

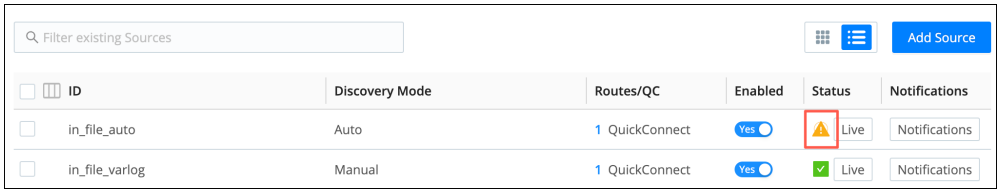


Source tile view

The colors have the following meanings for each Source and Destination:

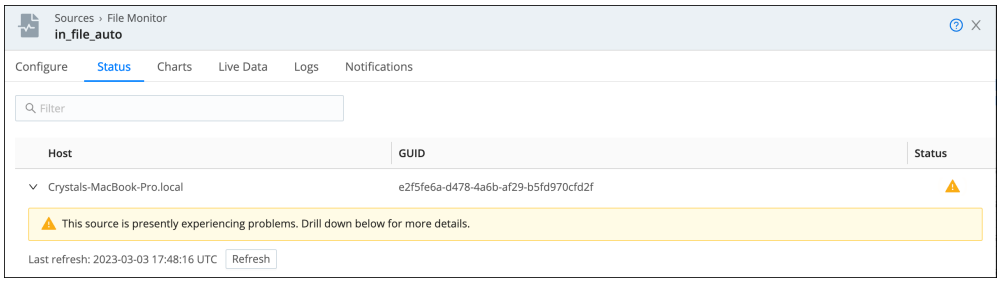
- **Green**: Working properly
- **Orange**: Experiencing problems
- **Red**: Experiencing severe problems
- **Gray**: Not enabled

To learn more about the health status of any Source or Destination, click its name to open the list of existing Sources or Destinations, and then click the Status icon for the one you want to examine.



Health status icon

In the modal that opens, click the Status icon again to see more details.



Health status details

When there are more than 100 Nodes in the Fleet, the **Status** column will not be visible in the list of Nodes. To view the status for an individual Node, expand it in the list. You can only expand one Node at a time. Refresh the list of Nodes and their status with the **Refresh** button.

Enabling Health Status

Health status must be enabled independently for each fleet in Fleet Settings. To enable health status for a Fleet:

1. Select the Fleet you want to manage.
2. On the **Manage** page, select the **Fleet Settings** tab.
3. Navigate to **General Settings** > **Limits** > **Metrics**.
4. Under **Metrics to send from Edge Nodes**, use the buttons to select or create a set of metrics.

If **Minimal** is selected, health status will not be enabled for the Fleet and health icons will appear white for all Sources and Destinations.

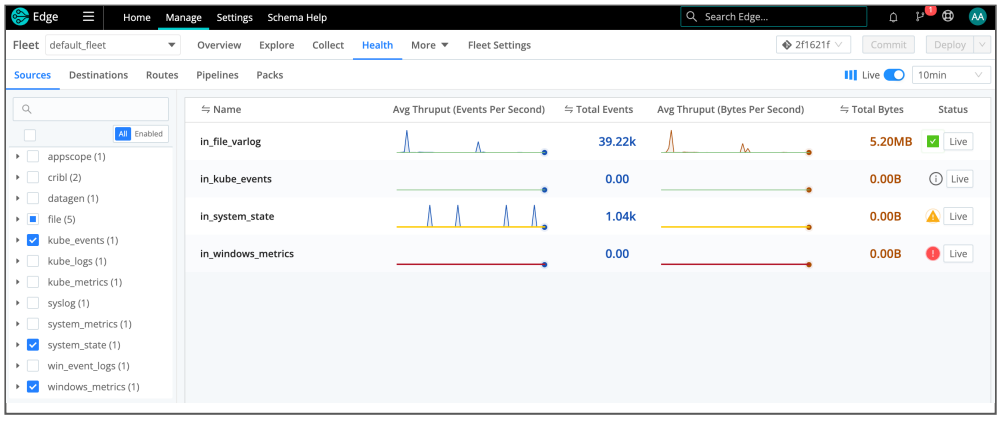
To learn more about these metrics options, see Controlling Metrics.

Health Status

You can isolate throughput for any of the following:

- Sources
- Destinations
- Routes
- Pipelines
- Packs
- Data Fields (Edge Nodes only)

For Fleets, go to the **Health** tab and click the tab you want. For Edge Nodes, go to the Edge Node's **Health** tab, and then select an option on the **Data** submenu.



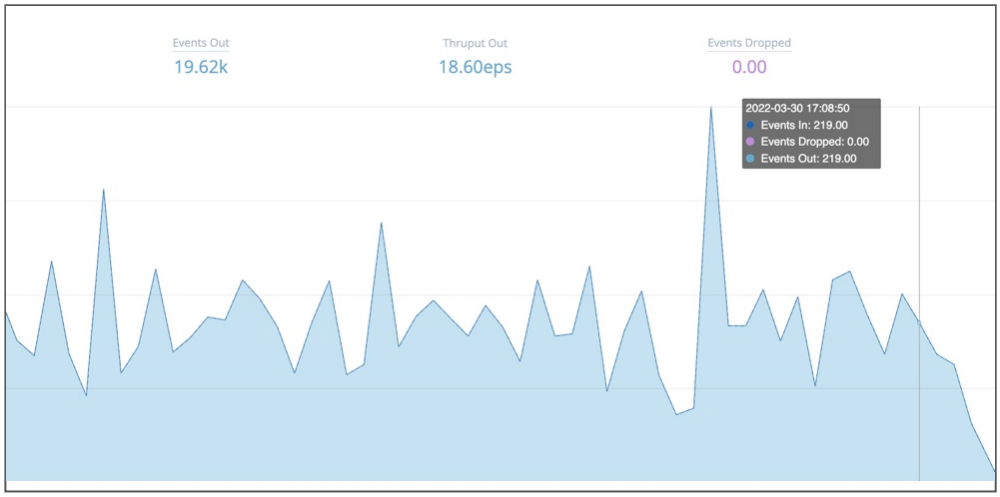
Manage > Health tab

Dense displays are condensed to sparklines for legibility.

The **Status** column displays the health of each resource with a colored icon. Click the icon to see more details. If a Fleet's **Health** tab is empty, health status might not be enabled because only minimal metrics are being sent. You can enable health status by going to **Fleet Settings** > **General Settings** > **Metrics** and selecting **Basic** or **All** under **Metrics to send from Edge Nodes**. See Controlling Metrics to learn more about these options.

Hover over the right edge to display Maximize buttons that you can click to zoom these up to detailed graphs.

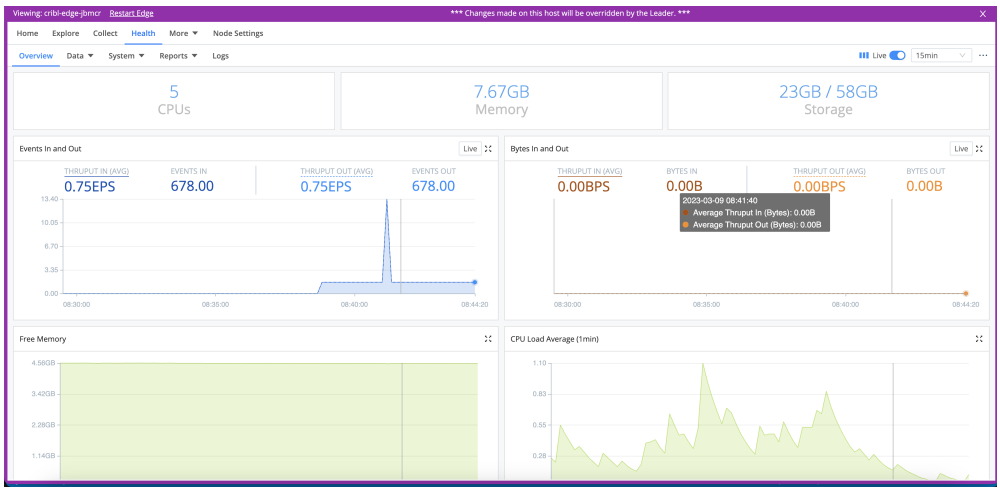
You can hover over an expanded graph fly-out to display further details.



Throughput details

Edge Node Health

To review the health status of an Edge Node, teleport into the Node. From the **Fleet** landing page, click the **List View** tab, and then click the **Edge Node GUID** link to teleport from the Leader into the Edge Node. From the top nav of the Edge Node, click the **Health** tab for monitoring details.



Metrics are displayed in the Cribl Edge UI in units of KB, MB, GB, and TB. At each level, these are multiples of `1024`.

The displayed **Storage** represents the amount of free storage remaining on the partition where Cribl Stream is installed. (This quantity might not represent the maximum storage available for the selected Edge Node or Fleet. Also, it does not calculate the system free space.)

Similarly, the **Free Memory** graph reflects only the operating system's `free` statistic, matching Linux's strict `free` definition by excluding `buff/cache` memory. So this graph indicates a lower value than the OS' `available` memory statistic – and it does not necessarily indicate that the OS is running out of memory to allocate.

Byte-related charts show the uncompressed amounts and rates of data processed over the selected time range:

- Events (total) in and out
- Events per second in and out
- Bytes (total) in and out
- Bytes per second in and out

We measure bytes in and out based on the size of `_raw`, if this field is present and is of type `string`. Otherwise, we use the size of read (uncompressed) data.

The displayed **CPU Load Average** is an average Edge Node Process, updated at 1-minute granularity. (It is not an average for the Edge Node as a whole.)

Monitoring data does not persist across Cribl Edge restarts. Keep this in mind before you restart the server.

System Monitoring

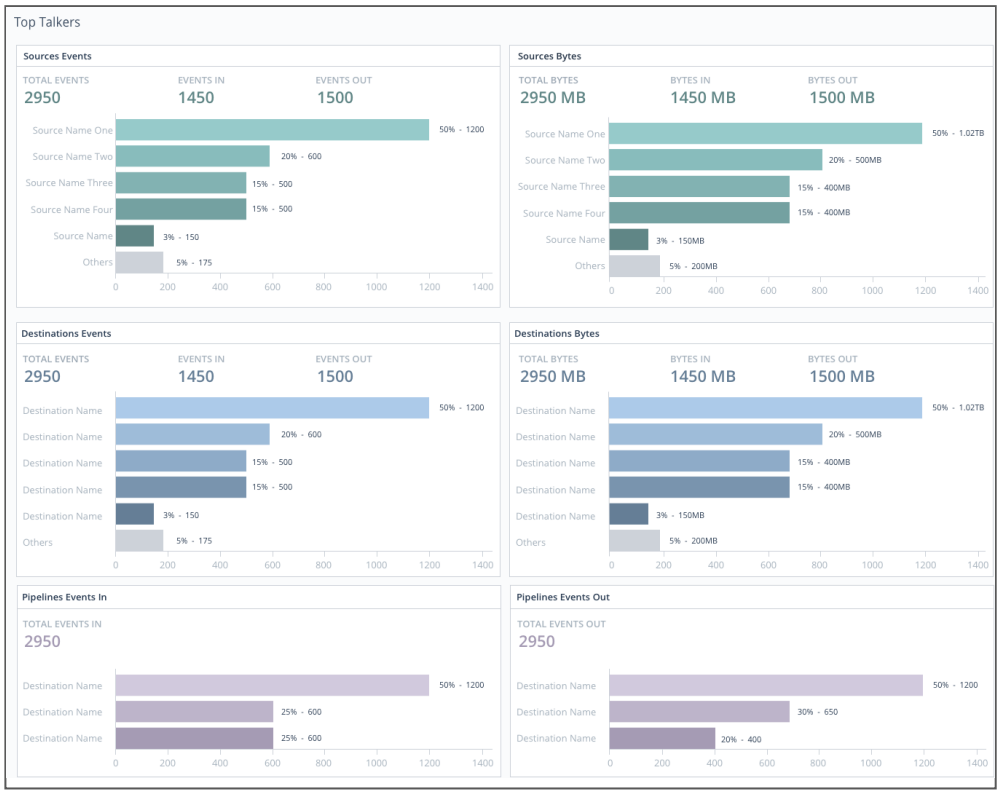
From the **Health** page's top nav, open the **System** submenu to isolate throughput for:

- Queues (Sources)
- Queues (Destinations)

For details, see Persistent Queues.

Reports/Top Talkers

Select **Health** > **Reports / Top Talkers** > **Top Talkers**, where you can examine your five highest-volume Sources, Destinations, Pipelines, Routes, and Packs. All components are ranked by events throughput. Sources and Destinations get separate rankings by bytes in and out, respectively.



Top Talkers

Internal Logs and Metrics

Select **Logs** from the **Health** page's top nav. Cribl Edge's internal logs and internal metrics provide comprehensive information about an instance's status/health, inputs, outputs, Pipelines, Routes, Functions, and traffic.

Health Endpoint

Query this endpoint on any instance to check the instance's health. (For details, see Querying the Health Endpoint.)

Types of Logs

Cribl Edge provides the following log types, by originating process:

- API Server Logs – These logs are emitted primarily by the `API/main` process. They correspond to the top-level `cribl.log` that shows up on the Diag page. These include telemetry/license-validation logs. Filesystem location: `$CRIBL_HOME/log/cribl.log`
- Edge Node Process(es) Logs – These logs are emitted by all the Edge Node Processes, and are very common on single-instance deployments and Edge Nodes. Filesystem location: `$CRIBL_HOME/log/worker/N/cribl.log`
- Fleet Logs – These logs are emitted by all processes that help a Leader Node configure Fleets. Filesystem location: `$CRIBL_HOME/log/group/GROUPNAME/cribl.log`

For details about generated log files, see Internal Logs. To work with logs in Cribl Edge's UI, see Search Internal Logs.

Cribl Edge rotates logs every 5 MB, keeping the most recent 5 logs. In a distributed deployment (Edge, all Edge Nodes forward their metrics to the Leader Node, which then consolidates them to provide a deployment-wide view.

Forward Logs and Metrics Externally

Cribl Edge supports forwarding internal logs and metrics to your preferred external monitoring solution. To make internal data available to send out, go to **More** > **Sources** and enable the Cribl Internal Source.

This will send internal logs and metrics down through Routes and Pipelines, just like another data source. Both logs and metrics will have a field called `source`, set to the value `cribl`, which you can use in Routes' filters.

Note that the only logs supported here are Edge Node Process logs (see Types of Logs above). You can, however, use a Script Collector to listen for API Server or Edge Node Fleet events.

For recommendations about useful Cribl metrics to monitor, see Internal Metrics.

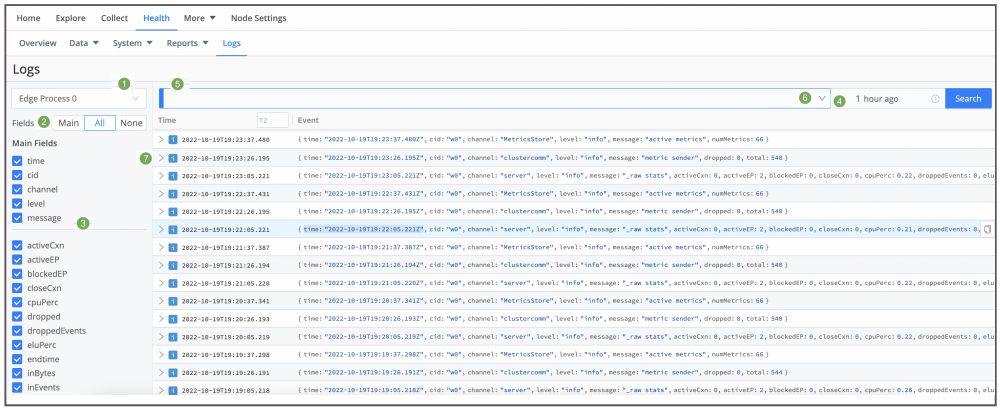
CriblMetrics Override

The **Disable field metrics** setting – in **Settings** (top nav) > **System > General Settings > Limits** - applies only to metrics sent to the Leader Node. When the **Cribl Internal** Source is enabled, Cribl Edge ignores this **Disable field metrics** setting, and full-fidelity data will flow down the Routes.

Search Internal Logs

Cribl Edge exists because logs are great and wonderful things! Using Cribl Edge's **Health** > **Logs** page, you can search all Cribl Edge's internal logs at once – from a single location, for both Leader and Edge Node. This enables you to query across all internal logs for strings of interest.

The labels on this screenshot highlight the key controls you can use (see the descriptions below):



Logs page (controls highlighted)

1. **Log file selector**: Choose the types of logs to view.
2. **Fields selector**: Click the **Main** | **All** | **None** toggles to quickly select or deselect multiple check boxes below. Beside these toggles, a Copy button enables you to copy field names to the clipboard in CSV format.
3. **Fields**: Select or deselect these check boxes to determine which columns are displayed in the Results pane at right. (The upper **Main Fields** group will contain data for *every* event; other fields might not display data for all events.)
4. **Time range selector**: Select a standard or custom range of log data to display. (The **Custom Time Range** pickers use your local time, even though the logs' timestamps are in UTC.)
5. **Search box**: To limit the displayed results, enter a JavaScript expression here. An expression must evaluate to `truthy` to return results. You can press **Shift+Enter** to insert a newline.

To modify the depth of information that is originally input to the Logs page, see Logging Settings.

6. Click the Search box's history arrow (right side) to retrieve recent queries.
7. The Results pane displays most-recent events first. Each event's icon is color-coded to match the event's severity level.

Click individual log events to unwrap an expanded view of their fields.

Logging Settings

On Cribl Edge's Settings pages, you can adjust the level (verbosity) of internal logging data processed, per logging channel. You can also redact fields in customized ways. In a distributed deployment, you manage each of these settings on the Edge Node, Fleet Settings, or main Settings for all Fleets.

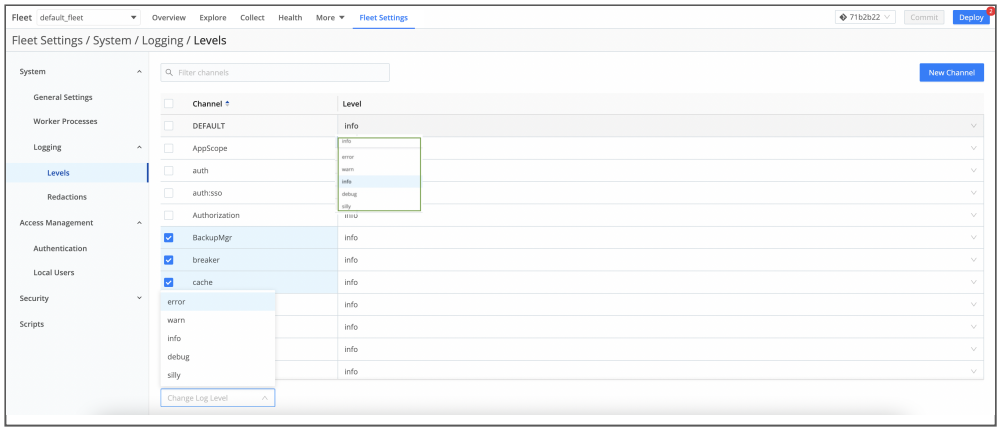
Change Logging Levels

To adjust logging levels:

- From Cribl Edge's top nav, select **Settings** > **System > Logging > Levels**.
- To configure log levels on a Fleet-level, click **Manage**, then select the Fleet you want to configure. Next, select **Fleet Settings** > **System > Logging > Levels**.
- To adjust settings on a a specific Edge Node, teleport into the Node, then select **Node Settings** > **System > Logging > Levels**.

On the resulting **Manage Logging Levels** page, you can:

- Modify one channel by clicking its **Level** column. In the resulting drop-down, you can set a verbosity level ranging from **error** up to **debug**. (Top of composite screenshot below.)
- Modify multiple channels by selecting their check boxes, then clicking the **Change log level** drop-down at the bottom of the page. (Bottom of composite screenshot below.) You can select all channels at once by clicking the top check box. You can search for channels at top right.



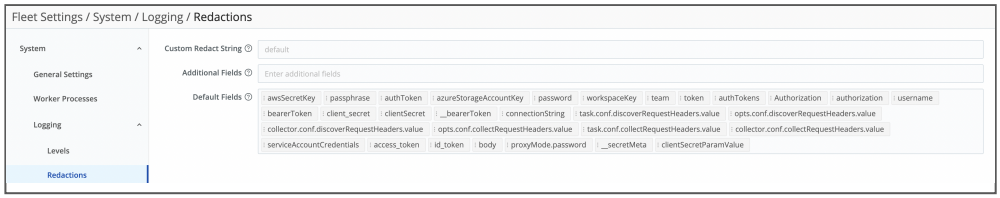
Manage Logging Levels page

The `silly` (ultra-verbose) logging level is available for all channels. However, it provides additional metrics information only for inputs.

Change Logging Redactions

On the **Redact Internal Log Fields** page, you can customize the redaction of sensitive, verbose, or just ugly data within Cribl Edge's internal logs. To access these settings:

- In a single-instance deployment, or for the Leader Node's own logs, select **Settings** (top nav) > **System** > **Logging** > **Redactions**.
- In a distributed deployment, first click **Manage**, then select the Fleet you want to configure. Next, select **Fleet Settings** > **System** > **Logging** > **Redactions**.



Redact Internal Log Fields page

It's easiest to understand the resulting **Redact Internal Log Fields** page's fields from bottom to top:

- **Default fields**: Cribl Edge always redacts these fields, and you can't modify this list to allow any of them through. However, you can use the two adjacent fields to define stricter redaction:
- **Additional fields**: Type or paste in the names of extra fields you want to redact. Use a tab or hard return to confirm each entry.
- **Custom redact string**: Unless this field is empty, it defines a literal string that will override Cribl Edge's default redaction pattern (explained below) on the affected fields.

Default Redact String

By default, Cribl Edge transforms this page's selected fields by applying the following redaction pattern:

- Echo the field value's first two characters.
- Replace all intermediate characters with a literal `...` ellipsis.
- Echo the value's last two characters.

Anything you enter in the **Custom redact string** field will override this default `??...??` pattern.

Querying the Health Endpoint

Each Cribl Edge instance exposes a `health` endpoint – typically used in conjunction with a Load Balancer – that you can use to make operational decisions.

Health Check Endpoint	Healthy Response	Cribl Edge Version
`curl http(s)://<host>:`	`{"status":"healthy"}`	Through 2.4.3

Health Check Endpoint	Healthy Response	Cribl Edge Version
`<port>/api/v1/health`		
`curl http(s)://<host>:<port>/api/v1/health`	`{"status":"healthy","startTime":1617814717110}` (see details below)	2.4.4 and later

Specifically, the `health` endpoint can return one of the following response codes:

- 200 – Healthy.
- 420 – Shutting down.
- 503 – HTTP engine reports server busy: too many concurrent connections (configurable).

In the above curl examples, `<port>` stands for the API port (by default, `9420`).

12.1. Internal Metrics

When sending Cribl Edge metrics to a metric system of analysis – such as InfluxDB, Splunk, or Elasticsearch – some metrics are particularly valuable. You can use these metrics to set up alerts when a Worker/Edge Node is having a problem, a Node is down, a Destination is down, a Source stops providing incoming data, etc.

Cribl Edge reports its internal metrics within the Cribl Edge UI, in the same way that it reports internal logs. Navigate to **Monitoring** > **Logs** (Cribl Stream), or to a Fleet's **Health** > **Sources** tab (Cribl Edge). To expose metrics for capture or routing, enable the **Cribl Internal** Source > **CriblMetrics** section. In Cribl Edge 4.4 and later, you can export metrics (and logs) to JSON files.

By default, Cribl Edge generates internal metrics every 2 seconds. To consume metrics at longer intervals, you can use or adapt the `cribl-metrics_rollup` Pipeline that ships with Cribl Stream. Attach it to your **Cribl Internal** Source as a pre-processing Pipeline. The Pipeline's **Rollup Metrics** Function has a default **Time Window** of 30 seconds, which you can adjust to a different granularity as needed.

You can also use our public endpoints to automate monitoring using your own external tools.

Counter-type metrics in Cribl Edge do not monotonically increase or decrease. They are reset at the end of each reporting period. Cribl Edge does not report counters when their value is 0. For example, if there aren't any Destinations reporting dropped events then the `total.dropped_events` metric won't be reported because its value would be 0.

Breaking Changes

- The `ci` and `co` dimensions are deprecated as of Cribl Edge 4.1.2. See Dimensions for their replacements.
- Disabled Sources no longer generate Cribl internal metrics as of Cribl Edge 4.2.

Total Throughput Metrics

Five important metrics below are prefixed with `total.` These power the top of Cribl Edge's **Monitoring** dashboard. The first two report on Sources, the remainder on Destinations.

- `total.in_bytes`
- `total.in_events`
- `total.out_events`
- `total.out_bytes`

- `total.dropped_events` – helpful for discovering situations such as: you've disabled a Destination without noticing.

Interpreting Total Metrics

These `total.` metrics' values could reflect Cribl Edge's health, but could also report low activity simply due to the Source system. E.g., logs from a store site will be low at low buying periods.

Also, despite the `total.` prefix, these metrics are each specific to the Worker/Edge Node Process that's generating them.

You can distinguish **unique** metrics by their `#input=<id>` dimension. For example, `total.in_events|#input=foo` would be one unique metric, while `total.in_events|#input=bar` would be another.

System Health Metrics

Five specific metrics are most valuable for monitoring system health. The first two are Cribl Edge composite metrics; the remaining three report on your hardware or VM infrastructure. Because the Cribl Internal Source does not export these metrics to Routes or Pipelines, you can obtain them only by using the REST endpoints listed on this page.

- `health.inputs`
- `health.outputs` – see the JSON Examples below for both `health.` metrics.
- `system.load_avg`
- `system.free_mem`
- `system.disk_used` – valuable if you know your disk size, especially for monitoring Persistent Queues. Here, a `0` value typically indicates that the disk-usage data provider has not yet provided the metric with data. (Getting the first value should take about one minute.)

All of the above metrics take these three values:

- `0` = green = healthy.
- `1` = yellow = warning.
- `2` = red = trouble.

Health Inputs/Outputs JSON Examples

The `health.inputs` metrics are reported per Source, and the `health.outputs` metrics per Destination. The `health.inputs` example below has two configured Sources, and two Cribl Stream-internal inputs. The

`health.outputs` example includes the built-in `devnull` Destination, and six user-configured Destinations.

Given all the `0` values here, everything is in good shape!

```
"health.inputs": [
  { "model": { "input": "http:http" }}, "val": 0},
  { "model": { "input": "cribl:CriblLogs" }}, "val": 0},
  { "model": { "input": "cribl:CriblMetrics" }}, "val": 0},
  { "model": { "input": "datagen:DatagenWeblog" }}, "val": 0 }
  ],
"health.outputs": [
  { "model": { "output": "devnull:devnull" }}, "val": 0},
  { "model": { "output": "router:MyOut1" }}, "val": 0},
  { "model": { "output": "tcpjson:MyTcpOut1" }}, "val": 0},
  { "model": { "output": "router:MyOut2" }}, "val": 0},
  { "model": { "output": "tcpjson:MyTcpOut2" }}, "val": 0},
  { "model": { "output": "router:MyOut3" }}, "val": 0},
  { "model": { "output": "router:MyOut4" }}, "val": 0 }
  ],
```

Worker/Edge Node Resource Metrics

As of Cribl Stream (LogStream) 2.4.4, the Cribl Internal Source reports two useful metrics on individual Worker/Edge Node Processes' resource usage:

- `system.cpu_perc` – CPU percentage usage.
- `system.mem_rss` – RAM usage.

Persistent Queue Metrics

These metrics are valuable for monitoring Persistent Queues' behavior:

- `pq.queue_size` – Total bytes in the queue.
- `pq.in_bytes` – Total bytes in the queue for the given **Time Window**.
- `pq.in_events` – Number of events in the queue for the given **Time Window**.
- `pq.out_bytes` – Total bytes flushed from the queue for the given **Time Window**.
- `pq.out_events` – Number of events flushed from the queue for the given **Time Window**.

These are aggregate metrics, but you can distinguish unique metrics per queue. On the Destination side, use the `#output=<id>` dimension. For example, `pq.out_events|#output=kafka` would be one unique metric; `pq.out_events|#output=camus` would be another.

Other Internal Metrics

The Cribl Internal Source emits other metrics that, when displayed in downstream dashboards, can be useful for understanding Cribl Edge's behavior and health. These include:

- `cribl.logstream.total.activeCxn` – Total active inbound TCP connections.
- `cribl.logstream.pipe.in_events` – Inbound events per Pipeline.
- `cribl.logstream.pipe.out_events` – Outbound events per Pipeline.
- `cribl.logstream.pipe.dropped_events` – Dropped events per Pipeline.
- `cribl.logstream.metrics_pool.num_metrics` – The total number of unique metrics that have been allocated into memory.
- `cribl.logstream.collector_cache.size` – Each Collector function (`default/cribl/collectors/<collector>/index.js`) is loaded/initialized only once per job, and then cached. This metric represents the current size of this cache.
- `cribl.logstream.cluster.metrics.sender.inflight` – Number of metric packets currently being sent from a Worker/Edge Node Process to the API Process, via IPC (interprocess communication).
- `cribl.logstream.backpressure.outputs` – Destinations experiencing backpressure, causing events to be either blocked or dropped.
- `cribl.logstream.blocked.outputs` – Blocked Destinations. (This metric is more restrictive than the one listed just above.)
- `cribl.logstream.pq.queue_size` – Current queue size, in bytes, per Worker Process.
- `cribl.logstream.host.in_bytes` – Inbound bytes from a given host (host is a characteristic of the data).
- `cribl.logstream.host.in_events` – Inbound events from a given host (host is a characteristic of the data).
- `cribl.logstream.host.out_bytes` – Outbound bytes from a given host (host is a characteristic of the data).
- `cribl.logstream.host.out_events` – Outbound events from a given host (host is a characteristic of the data).
- `cribl.logstream.index.in_bytes` – Inbound bytes per index.
- `cribl.logstream.index.in_events` – Inbound events per index.

- `cribl.logstream.index.out_bytes` – Outbound bytes per index.
- `cribl.logstream.index.out_events` – Outbound events per index.
- `cribl.logstream.route.in_bytes` – Inbound bytes per Route.
- `cribl.logstream.route.in_events` – Inbound events per Route.
- `cribl.logstream.route.out_bytes` – Outbound bytes per Route.
- `cribl.logstream.route.out_events` – Outbound events per Route.
- `cribl.logstream.source.in_bytes` – Inbound bytes per Source.
- `cribl.logstream.source.in_events` – Inbound events per Source.
- `cribl.logstream.source.out_bytes` – Outbound bytes per Source.
- `cribl.logstream.source.out_events` – Outbound events per Source.
- `cribl.logstream.sourcetype.in_bytes` – Inbound bytes per sourcetype.
- `cribl.logstream.sourcetype.in_events` – Inbound events per sourcetype.
- `cribl.logstream.sourcetype.out_bytes` – Outbound bytes per sourcetype.
- `cribl.logstream.sourcetype.out_events` – Outbound events per sourcetype.
- `nodecluster.primary.messages` – Number of IPC (interprocess communication) calls made by a Worker/Edge Node Process to the API Process.

Dimensions

Cribl Edge internal metrics have extra dimensions. These include:

- `cribl_wp` – Identifies the Worker/Edge Node Process that processed the event.
- `event_host` – Machine's hostname from which the event was made.
- `event_source` – Set as `cribl`.
- `group` – Name of Cribl Fleet from which the event was made.
- `input` – Entire **Input ID**, including the prefix, from which the event was made.
- `output` – Entire **Output ID**, including the prefix, from which the event was made.
- `ci` – Deprecated; use `input` instead. Cribl plans to drop support for this dimension after version 4.x.
- `co` – Deprecated; use `output` instead. Cribl plans to drop support for this dimension after version 4.x.

The `ci` and `co` (duplicate) dimensions are deprecated as of Cribl Edge 4.1.2, and will be removed in a future version. Where any of your code relies on these dimensions, Cribl recommends that you

> promptly substitute their respective replacements, `input` and `output`. This applies to relevant Cribl Functions and Pipelines, dashboards in downstream services, etc.

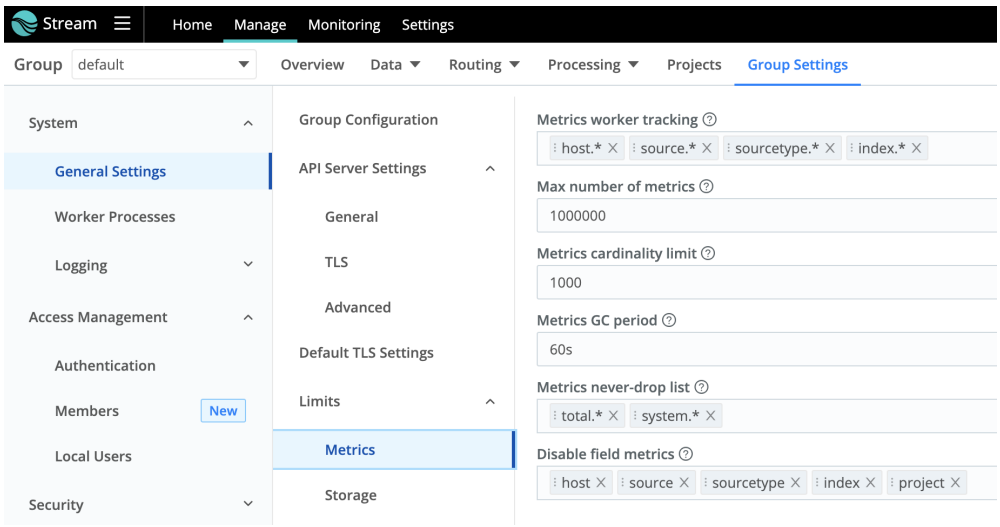
Controlling Metrics

You can control the volume of internal metrics that Cribl Edge emits, by adjusting Metrics Limits. You access these as follows, per deployment type:

- Single-instance: **Settings** > **System** > **General Settings** > **Limits** > **Metrics**.
- Distributed Leader: **Settings** > **Global Settings** > **System** > **General Settings** > **Limits** > **Metrics**.
- Distributed Fleets: [Select a Fleet] > **Fleet Settings** > **System** > **General Settings** > **Limits** > **Metrics**.

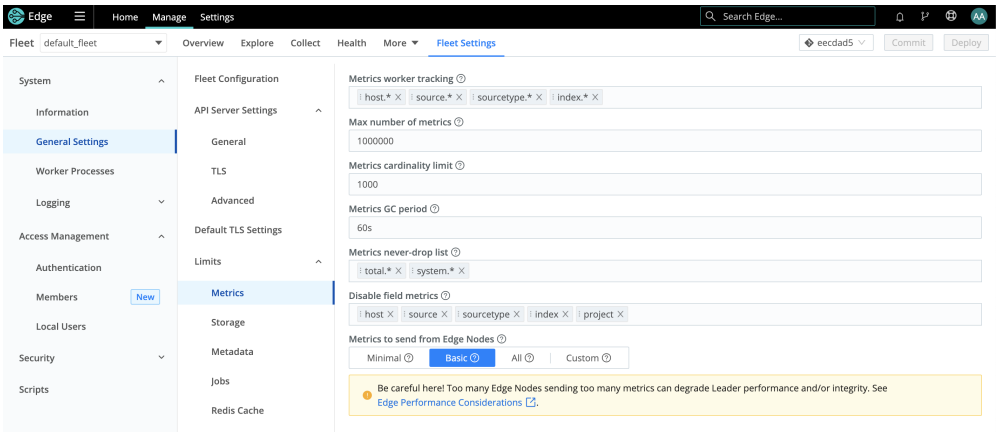
Here, you can adjust the maximum number of metric series allowed, metrics' cardinality, metrics to always include (**Metrics never-drop list**), metrics to always exclude in order to optimize performance (**Disable field metrics**), and other options.

Here's how these controls look in Cribl Stream:



Metrics settings for Cribl Stream Worker Groups

They're somewhat different in Cribl Edge – see the next section for an explanation:



Metrics settings for Cribl Edge Fleets

Controlling Metrics in Cribl Edge

Cribl Edge lets you choose what metrics each Fleet sends to its Leader Node. Your goal should be to collect the metrics most important to you, within the limits of the Leader's capacity to process them. This is a matter of balancing Fleet cardinality and Leader performance, as explained in the Deployment Planning topic.

Use the **Metrics to send from Edge Nodes** buttons to select one of the metrics sets listed in the table below.

Set	Purpose	List of Metrics Sent
Minimal	Send metrics for events and bytes in and out. Recommended for high-cardinality deployments.	Sent as aggregate totals per Fleet: `total.in_events`, `total.out_events`, `total.in_bytes`, `total.out_bytes` excluding metrics that are coming from Sources or Destinations directly.

Basic (default)	Send metrics required to display all monitoring data available in the Edge UI. Contains all the metrics in the Minimal set, and more.	Sent as aggregate totals per Source/Destination:
		
		`total.in_events`, `total.out_events`, `total.in_bytes`, `total.out_bytes`.
		
		Also, Sent as aggregate totals per Source:
		
		`health.inputs`
		
		Sent as aggregate totals per Destination:
		
		`health.outputs`
		
		Sent as aggregate totals per Route:
		
		`route.in_events`, `route.in_bytes`, `route.out_events`, `route.out_bytes`,

Set	Purpose	List of Metrics Sent

		`route.dropped_events`
		
		Sent as aggregate totals per Pipeline:
		
		`pipe.in_events`, `pipe.out_events`, `pipe.dropped_events`
		
		Sent as aggregate totals per Pack:
		
		`pack.in_events`, `pack.out_events`, `pack.err_events`, `pack.dropped_events`

All	Send all metrics with no filtering.	All metrics without filtering, including **Basic**, plus Worker/Edge Node Resource Metrics, Persistent Queue Metrics, and Other Internal Metrics.
Custom	Send a customized set of metrics.	All metrics that match a JavaScript expression that you define, plus the metrics in the **Minimal** set.

Every Edge Fleet always sends the **Minimal** metrics; selecting one of the other options only adds more metrics.

Choosing the **Minimal** metrics set has the following effects in the Edge UI:

- Only one UI page, **Manage** > **Overview** > **Monitor**, will be fully populated. For the metrics that the **Minimal** set excludes, other pages will show zero values. This applies to the **Manage** > **Health** pages for **Sources**, **Destinations**, **Routes**, **Pipelines**, and **Packs**.
- In the **Manage** > **More** pages for **Sources** and **Destinations**, the health indicators in the tile and list views do not function when you select the **Minimal** set. Rather than red, green, or gray, they'll always show white, because the relevant metrics, `health.inputs` and `health.outputs`, will not be sent to the Leader.

The simplest form of a **Custom** set just specifies additional desired metrics by name, using an expression of the form: `name === "metric.name"`. Among the most useful metrics to consider adding are `system.load_avg`, `system.free_mem`, `system.disk_used`, `system.cpu_perc`, and `system.mem_rss`.

Other Metrics Endpoints

Below is basic information on using the `/system/metrics` endpoint, the `/system/info` endpoint, and the `cribl_wp` dimension.

`/system/metrics` Endpoint

`/system/metrics` is Cribl Edge's primary public metrics endpoint, which returns most internal metrics. Note that many of these retrieved metrics report configuration only, not runtime behavior. For details, see our API Docs.

`/system/info` Endpoint

`/system/info` generates the JSON displayed in the Cribl Edge UI at **Settings** > **Global Settings** > **Diagnostics** > **System Info**. Its two most useful properties are `loadavg` and `memory`.

`loadavg` Example

```
"loadavg": [1.39599609375, 1.22265625, 1.31494140625],
```

This property is an array containing the 1-, 5-, and 15-minute load averages at the UNIX OS level. (On Windows, the return value is always `[0, 0, 0]`.) For details, see the Node.js os.loadavg() documentation.

`memory` Example

```
"memory": { "free": 936206336, "total": 16672968704 },
```

Divide `total` / `free` to monitor memory pressure. If the result exceeds 90%, this indicates a risky situation: you're running out of memory.

`cpus` Alternative

The `cpus` metric returns an array of CPU/memory key-value pairs. This provides an alternative way to understand CPU utilization, but it requires you to query all your CPUs individually.

12.2. Internal Logs

Distributed deployments emit a larger set of logs than single-instance deployments. We'll describe the distributed set first.

You can display and export all internal logs from the UI at **Monitoring** > **Logs** (Stream) or **Manage** > **Logs** (Edge). Logs' persistence depends on event volume, not time – for details, see Log Rotation and Retention.

Currently, most logs listed on this page are available only in customer-managed deployments. Cribl.Cloud currently supports only Edge Node logs, and only on hybrid (not Cribl-managed) Workers.

Leader Node Logs (Distributed)

The `API/main` process emits the following logs into the Leader Node's `$CRIBL_HOME/log/` directory.

Logfile Name	Description	Equivalent on Logs page
`cribl.log`	Principal log in Cribl Edge. Includes telemetry/license-validation logs. Corresponds to top-level `cribl.log` on Diag page.	**Leader** > **API Process**
`access.log`	API calls, e.g., `GET /api/v1/version/info`.	**Leader** > **Access**
`audit.log`	Actions pertaining to files, e.g., `create`, `update`, `commit`, `deploy`, `delete`.	**Leader** > **Audit**
`notifications.log`	Messages that appear in the notification list in the UI.	**Leader** > **Notifications**
`ui-access.log`	Interactions with different UI components described as URLs, e.g., `/settings/apidocs`, `/dashboard/logs`.	**Leader** > **UI Access**

The `API/main` process emits the following service logs into the Leader Node's `$CRIBL_HOME/log/service/` directory. Each service includes a `cribl.log` file that logs the service's internal telemetry and an `access.log` file that logs which API calls the service has handled.

SERVICE NAME	DESCRIPTION	EQUIVALENT ON Logs PAGE
Connections Service	Handles all worker connections and communication, including heartbeats, bundle deploys, teleporting, restarting, etc. Workers are assigned to connection processes using a round-robin algorithm.	**Leader** > **Connections Service**
Lease Renewal Service	Handles lease renewal for the primary Leader Node.	**Leader** > **Lease Renewal Service**
Metrics Service	Handles in-memory metrics, merging of incoming packets, metrics persistence and rehydration, and UI queries for metrics.	**Leader** > **Metrics Service**
Notifications Service	Triggers notifications based on its configuration.	**Leader** > **Notifications Service**

The Config Helper process for each Worker Group/Fleet emits the following log in `$CRIBL_HOME/log/group/GROUPNAME`.

Logfile Name	Description	Equivalent on Logs page
`cribl.log`	Messages about config maintenance, previews, etc.	**GROUPNAME** > **Config helper**

Edge Node Logs (Distributed)

The API Process emits the following log in `$CRIBL_HOME/log/`.

Logfile Name	Description	Equivalent on Logs page
`cribl.log`	Messages about the Worker/Edge Node communicating with the Leader Node (i.e., with its API Process) and other API requests, e.g., sending metrics, reaping job artifacts.	**GROUPNAME** > **Worker:HOSTNAME** > **API Process**

Each Worker Process emits the following logs in `$CRIBL_HOME/log/worker/N/`, where `N` is the Worker/Edge Node Process ID. (The `metrics.log` file is written only when HTTP-based Destinations exist.)

Logfile Name	Description	Equivalent on Logs page
`cribl.log`	Messages about the Worker/Edge Node processing data.	**GROUPNAME** > **Worker:HOSTNAME** > **Worker Process N**

Logfile Name	Description	Equivalent on Logs page
`metrics.log`	Messages about the Worker/Edge Node's outbound HTTP request statistics.	**GROUPNAME** > **Worker:HOSTNAME** > **Worker Process N**

For convenience, the UI aggregates the Worker/Edge Node Process logs as follows.

Logfile Name	Description	Equivalent on Logs page
N/A	Aggregation of all the `Worker Process N` logs and the API Process log.	**GROUPNAME** > **WORKER_NAME**

The Edge Node logs listed above are currently available only on hybrid (not Cribl-managed) Workers. The single-instance logs listed below are not relevant to Cribl.Cloud.

Single-Instance Logs

The `API/main` process emits the same logs as it does for a distributed deployment, in `$CRIBL_HOME/log/`:

- `cribl.log`
- `access.log`
- `audit.log`
- `notifications.log`
- `ui-access.log`

Each Worker/Edge Node Process emits the following logs in `$CRIBL_HOME/log/worker/N/`, where `N` is the Worker/Edge Node Process ID. (The `metrics.log` file is written only when HTTP-based Destinations exist.)

Logfile Name	Description	Equivalent on Logs page
`cribl.log`	Messages about the Worker/Edge Node processing data.	**GROUPNAME** > **Worker:HOSTNAME** > **Worker Process N**
`metrics.log`	Messages about the Worker/Edge Node's outbound HTTP request statistics.	**GROUPNAME** > **Worker:HOSTNAME** > **Worker Process N**

`_raw stats` Event Fields

Each Worker/Edge Node Process logs this information at a 1-minute frequency. So each event's scope covers only that Worker/Edge Node Process, over a 1-minute time span defined by the `startTime` and `endTime` fields.

Sample Event


```
{"time":"2022-11-17T16:54:05.349Z","cid":"w0","channel":"server","level":"info","me
```


Field Descriptions

Field	Description
`abortCxn`	Number of TCP connections that were aborted.
`activeCxn`	Number of TCP connections newly opened at the time the `_raw` stats are logged. (This is a gauge when exported in internal metrics, and can otherwise be ignored as an instantaneous measurement. Only some application protocols count toward this; e.g., any HTTP-based Source does not count.)
`activeEP`	Number of currently active event processors (EPs). EPs are used to process events through Breakers and Pipelines as the events are received from Sources and sent to destinations. EPs are typically created per TCP connection (such as for HTTP).
`blockedEP`	Number of currently blocked event processors (caused by blocking Destinations).
`closeCxn`	Number of TCP connections that were closed.
`cpuPerc`	CPU utilization from the combined user and system activity over the last 60 seconds.
`droppedEvents`	This is equivalent to the `total.dropped_events` metric. Drops can occur from Pipeline Functions, from Destination Backpressure, or from other errors. Any event not sent to a Destination is considered dropped.
`eluPerc`	Event loop utilization. Represents the percentage of time over the last 60 seconds that the NodeJS runtime spent processing events within its event loop.
`endTime`	The end of the timespan represented by these metrics. (Will always be 60 seconds after `startTime`.)
`inBytes`	Number of bytes received from all Sources (based only off `_raw`).
`inEvents`	Number of events received from all inputs after Event Breakers are applied. This can be larger than `outEvents` if events are dropped via Drop, Aggregation, Suppression, Sampling, or similar Functions.

Field	Description
`mem.buffers`	Memory allocated for ArrayBuffers and SharedArrayBuffers.
`mem.ext`	External section of process memory, in MB.
`mem.heap`	Used heap section of process memory, in MB.
`mem.heapTotal`	Total heap section of process memory, in MB.
`mem.rss`	Resident set size section of process memory, in MB.
`openCxn`	Same as `activeCxn`, but tracked as a counter rather than a gauge. So `openCxn` will show all connections newly opened each minute, and is more accurate than using `activeCxn`.
`outBytes`	Number of bytes sent to all Destinations (based only off `_raw`).
`outEvents`	Number of events sent out to all Destinations. This can be larger than `inEvents` due to creating event clones or entirely new unique events (such as when using the Aggregation Function).
`pqInBytes`	Number of bytes that were written to Persistent Queues, across all Destinations.
`pqInEvents`	Number of events that were written to Persistent Queues, across all Destinations.
`pqOutBytes`	Number of bytes that were flushed from Persistent Queues, across all Destinations.
`pqOutEvents`	Number of events that were flushed from Persistent Queues, across all Destinations.
`pqTotalBytes`	Amount of data currently stored in Persistent Queues, across all Destinations.
`rejectCxn`	Number of TCP connections that were rejected.
`startTime`	The beginning of the timespan represented by these metrics.
`tasksCompleted`	The number of tasks the process has started and completed for all collection jobs for which it was executing tasks.
`tasksStarted`	The number of tasks the process started for all collection jobs for which it was executing tasks.

13. Working With Data

13.1. Event Model

All data processing in Cribl Edge is based on discrete data entities commonly known as **events**. An event is defined as a collection of key-value pairs (fields). Some Sources deliver events directly, while others might deliver bytestreams that need to be broken up by Event Breakers. Events travel from a Source through Pipelines' Functions, and on to Destinations.

The internal representation of a Cribl Edge event is as follows:

Event Model

```
{
  "_raw": "<body of non-JSON parse-able event>",
  "_time": "<timestamp in UNIX epoch format>",
  "__inputId": "<Id/Name of Source that delivered the event>",
  "__other1": "<Internal field1>",
  "__other2": "<Internal field2>",
  "__otherN": "<Internal fieldN>",
  "key1": "<value1>",
  "key2": "<value2>",
  "keyN": "<valueN>",
  "...": "..."
}
```

Some notes about these representative fields:

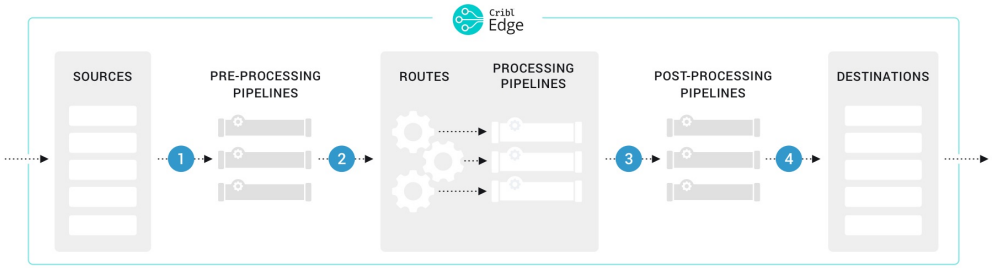
- Fields that start with a double-underscore are known as **internal fields**, and each Source can add one or many to each event. For example, Syslog adds both a `__inputId` and a `__srcIpPort` field. Internal fields are used only within Cribl Edge, and are not passed down to Destinations.
- Upon arrival from a Source, if an event cannot be JSON-parsed, all of its content will be assigned to `_raw`.
- If a timestamp is not configured to be extracted, the current time (in UNIX epoch format) will be assigned to `_time`.
- Cribl reserves the right to change all internal fields that are not documented under Sources or Destinations.

Using Capture

One way to see what an event looks like as it travels through the system is to use the **Capture** feature. While in Preview (right pane):

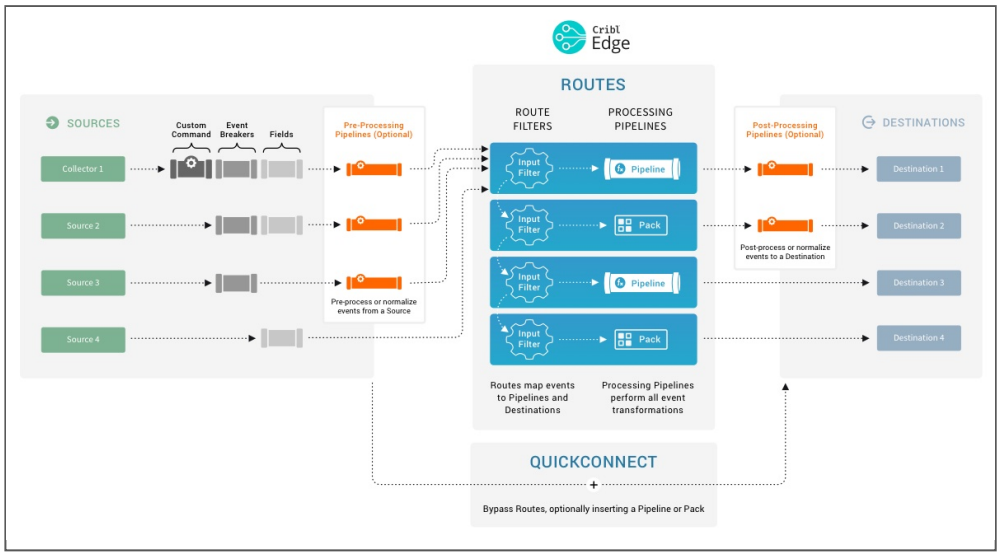
1. Click **Start a Capture**.
2. In the resulting modal, enter a **Filter expression** to narrow down the events of interest.
3. Click **Capture...** and (optionally) change the default Time and/or Event limits.
4. Select the desired **Where to capture** option. There are four options:

- **1. Before the pre-processing Pipeline** – Capture events right after they're delivered by the respective Input.
- **2. Before the Routes** – Capture events right after the pre-processing Pipeline, before they go down the Routes.
- **3. Before the post-processing Pipeline** – Capture events right after the Processing Pipeline that actually handled them, before any post-processing Pipeline.
- **4. Before the Destination** – Capture events right after the post-processing Pipeline, before they go out to the configured Destination.



13.2. Event Processing Order

The expanded schematic below shows how all events in the Cribl Edge ecosystem are processed linearly, from left to right.



Cribl Edge in great detail

Here are the stages of event processing:

1. Sources: Data arrives from your choice of external providers. (Cribl Edge supports Splunk, HTTP/S, Elastic Beats, Amazon Kinesis/S3/SQS, Kafka, TCP raw or JSON, and many others.)
2. Custom command: Optionally, you can pass this input's data to an external command before the data continues downstream. This external command will consume the data via `stdin`, will process it and send its output via `stdout`.
3. Event Breakers can, optionally, break up incoming bytestreams into discrete events. (Note that because Event Breakers precede tokens, Breakers cannot see or act on tokens.)
4. Auth tokens are applied at this point on Splunk HEC Sources. (Details here.)
5. Fields/Metadata: Optionally, you can add these enrichments to each incoming event. You add fields by specifying key-value pairs, per Source, in a format similar to Cribl Edge's Eval Function. Each key defines a field name, and each value is a JavaScript expression (or constant) used to compute the field's value.
6. Auth tokens are applied at this point on HTTP-based Sources other than Splunk HEC (e.g., Elasticsearch API Sources).
7. Pre-processing Pipeline: Optionally, you can use a single Pipeline to condition (normalize) data from this input before the data reaches the Routes.
8. Routes map incoming events to Processing Pipelines and Destinations. A Route can accept data from multiple Sources, but each Route can be associated with only one Pipeline and one Destination.

9. Processing Pipelines perform all event transformations. Within a Pipeline, you define these transformations as a linear series of Functions. A Function is an atomic piece of JavaScript code invoked on each event.
10. Post-processing Pipeline: Optionally, you can append a Pipeline to condition (normalize) data from each Processing Pipeline before the data reaches its Destination.
11. Destinations: Each Route/Pipeline combination forwards processed data to your choice of streaming or storage Destination. (Cribl Edge supports Splunk, Syslog, Elastic, Kafka/Confluent, Amazon S3, Filesystem/NFS, and many other options.)

Pipelines Everywhere

All Pipelines have the same basic internal structure – they're a series of Functions. The three Pipeline types identified above differ only in their position in the system.

13.3. Routes

Before incoming events are transformed by a processing Pipeline, Cribl Edge uses a set of filters to first select a **subset** of events to deliver to the correct Pipeline. This selection is normally made via Routes.

Don't Need Routes?

Routes are designed to filter, clone, and cascade incoming data across a related set of Pipelines and Destinations. If all you need are independent connections that link parallel Source/Destination pairs, you can use Cribl Edge's QuickConnect rapid visual configuration tool as an alternative to Routes.

Accessing Routes

Select **Routing** > **Data Routes** from Cribl Edge's global top nav (single-instance deployments) or from a Worker Group's/Fleet's top nav (distributed deployments). To configure a new Route, click **Add Route**.

How Do Routes Work

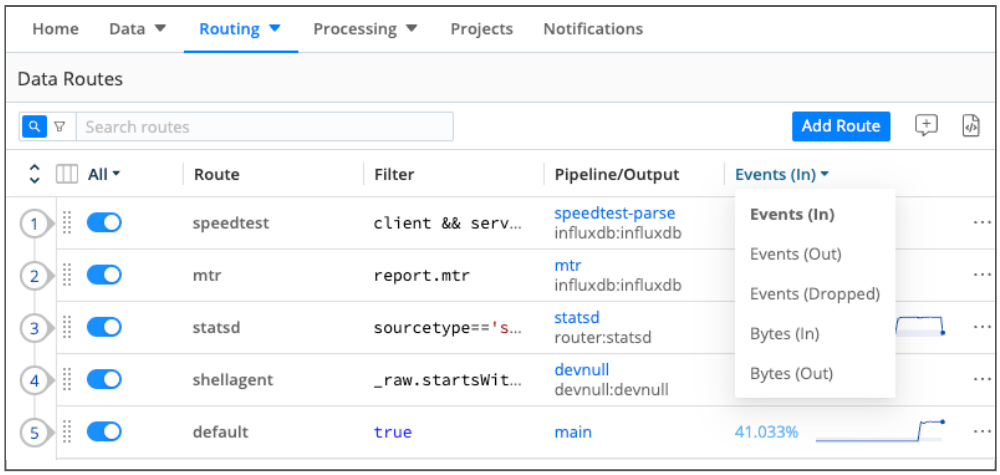
Routes apply filter expressions on incoming events to send matching results to the appropriate Pipeline. Filters are JavaScript-syntax–compatible expressions that are configured with each Route. Examples are:

- `true`
- `source=='foo.log' && fieldA=='bar'`

There can be multiple Routes in a Cribl Edge deployment, but each Route can be associated with only **one processing** Pipeline. (Sources and Destinations can also have their own attached pre-/post-processing Pipelines.)

Each Route must have a unique name. The **Route Name** field is case-sensitive, so entering the same characters with different capitalization will create a Route with a separate name.

Routes are evaluated in their display order, top->down. The stats shown in the **Bytes**/**Events** (toggle) column are for the most-recent 15 minutes.



Routes and bytes

In the example above, incoming events will be evaluated first against the Route named **speedtest**, then against **mtr**, then against **statsd**, and so on. At the end, the **default** Route serves as a catch-all for any event that does not match any of the other Routes.

Above, note the **Events In** drop-down menu to toggle between displaying Events versus Bytes, and to display In, Out, or Dropped.

Displayed percentages of Events and Bytes In and Out are by comparison to throughput or filtering across all Routes. For example, if you have four Routes that share an identical filter, each Route's Events In might show approximately `25%` of the total inbound traffic across all Routes.

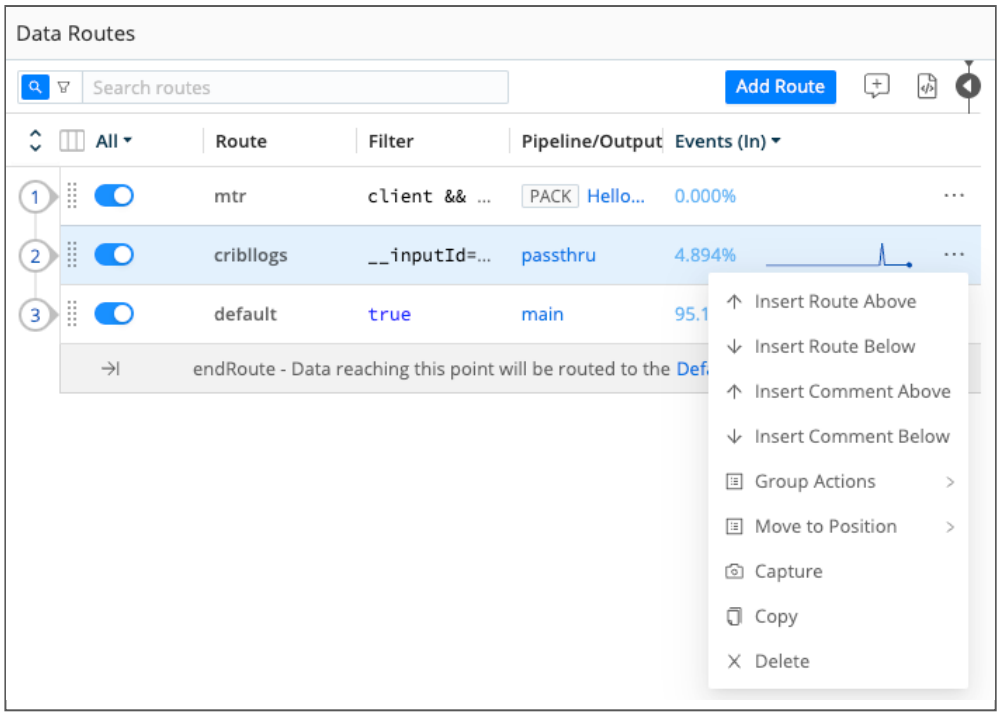
Managing the Routes Page

To apply a Route before another, simply drag it vertically. Use the toggles to turn Routes `On/Off` inline, as necessary, to facilitate development and debugging.

You can press the] (right-bracket) shortcut key to toggle between the Preview pane and the expanded Routes display shown above. (This works when no field has focus.)

Click a Route's Options (•••) menu to display multiple options:

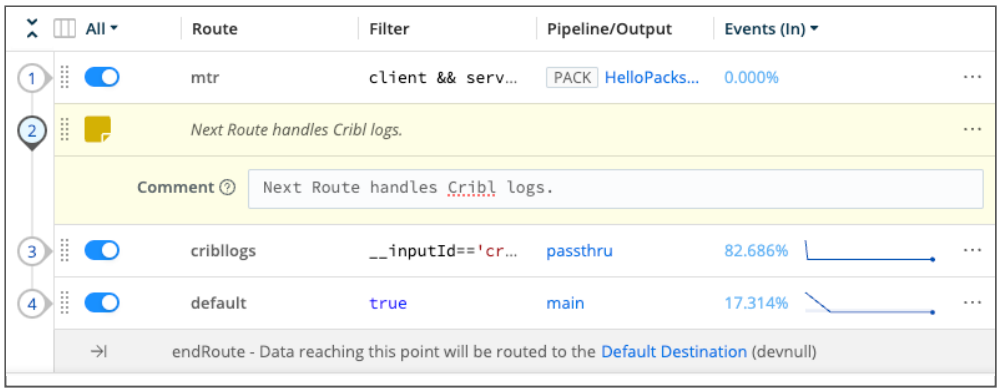
- Insert, group, move, copy, or delete Routes.
- Insert comments above or below Routes.
- Capture sample data through a selected Route.



Route > Options menu

Copying a Route displays a confirmation message and adds a Paste button next to **Add Route**. Pasting creates an exact duplicate of the Route, with a warning indicator to change its duplicate name.

Comments work like a Pipeline's Comment Functions: You can use them to describe Routes' purposes and/or interactions.



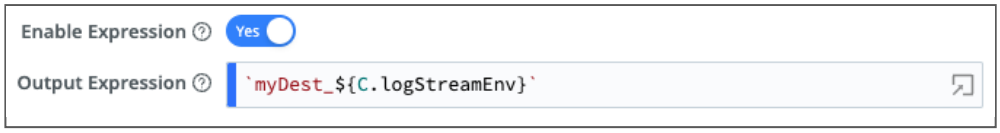
Comments make the Routing table self-documenting

Output Destination

You can configure each Route with an **Output** that defines a Destination to send events to, after they're processed by the Pipeline.

If you toggle **Enable expression** to `Yes`, the **Output** field changes to an **Output expression** field. Here, you can enter a JavaScript expression that Cribl Edge will evaluate as the name of the Destination. Sample

expression format: `` `myDest_${C.logStreamEnv}` ``. (This evaluation happens at Route construction time, not per event.)



Output expression enabled

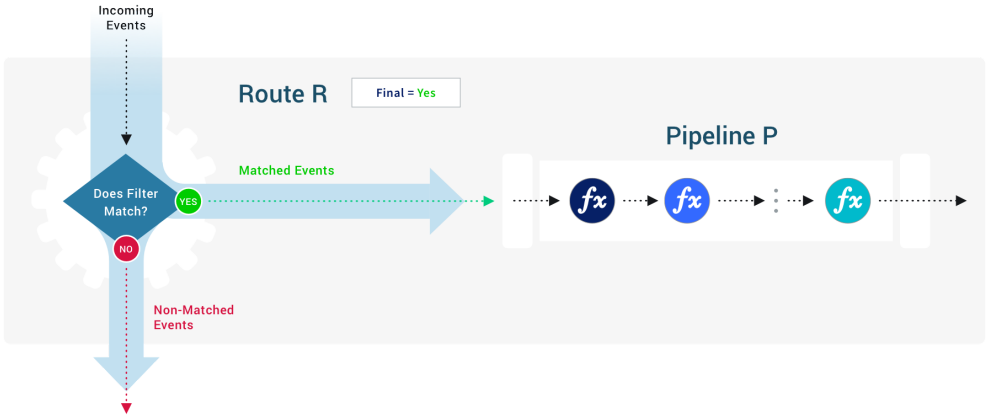
Expressions that match no Destination name will silently fail.

On Routes within Packs, neither **Output** control is available, because Packs cannot specify Destinations.

The `Final` Toggle

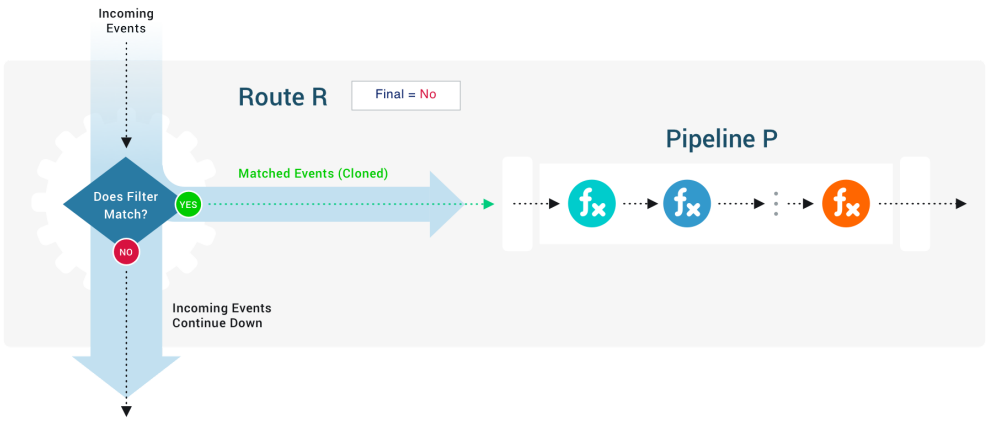
The `Final` toggle in Route settings controls what happens to an event after a Route-Pipeline pair matches it.

When `Final` is toggled to `Yes` (default), the Route "consumes" an event, meaning the event will not pass down to any other Route below.



The Route is Final and non-matched events stop here

When `Final` is toggled to `No`, the event will be processed by this Route, **and** then passed on to the next Route below.



The Route isn't Final and all events pass further on

When you toggle `Final` to `No`, the event sent to the current Route technically becomes a *clone* of the original event. The **Add Clone** button lets you add a fields – names and values – to the cloned events sent to this Route's Pipeline.

Non-final Route: add clone fields

Event Cloning Strategy

Depending on your cloning needs, you might want to follow a **most-specific first** or a **most-general first** processing strategy. The general goal is to minimize the number of filters/Routes an event gets evaluated against. For example:

If you don't need to clone any events (that is, `Final` is toggled to `Yes` everywhere), then it makes sense to start with the broadest expression at the top. This will consume as many events as early as possible.

If you need to clone a narrow set of events, then it might make sense to do that upfront. Then, follow it with a Route that consumes those clones immediately after.

The `endRoute` Bumper

The `endRoute` bumper appears at the bottom of the Routing table. This is a reminder that, if **no** Route in the table has the `Final` flag enabled, events will continue to Cribl Edge's configured Default Destination.

endRoute warning and link

This is a backstop to ensure data flow. However, if that configured default is also configured as the **Output** of a Route higher in the table, duplicate events will reach that Destination.

You can correct this either by setting a Route to `Final`, or by changing the Default Destination. The bumper provides a link to the Default Destination's config, and identifies the currently configured default in parentheses.

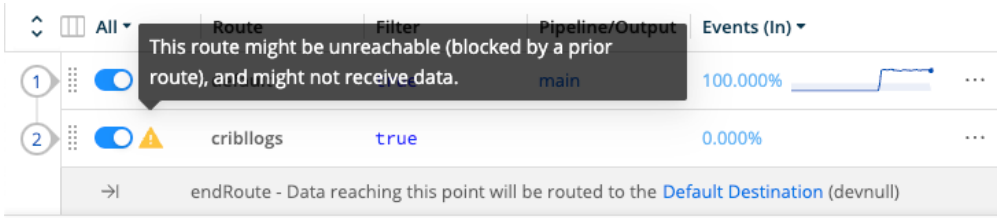
Route Groups

A Route group is a collection of consecutive Routes that can be moved up and down the Route stack together. Groups help with managing long lists of Routes. They are a UI visualization only: While Routes are in a group, those Routes maintain their global position order.

Route groups work much like Function groups, offering similar UI controls and drag-and-drop options.

Unreachable Routes

Routes display an "unreachable" warning indicator (orange triangle) when data can't reach them.



Unreachable Route warning, on hover

This condition will occur when, with your current configuration, any Route higher in the stack matches **all** three of these conditions:

- Previous Route is enabled (toggle is set to `On`).
- Previous Route is final (**Final** toggle is set to `Yes`).

- Previous Route's **Filter** expression evaluates to true, (e.g., `true`, `1 === 1`, etc.).

Note that the third condition above can be triggered intermittently by a randomizing method like `Math.random()`. This might be included in a previous Route's own Filter expression, or in a Pipeline Function (such as one configured for random data sampling).

Routing with Output Router

Output Router Destinations offer another way to route data. These function as meta-Destinations, in that they allow you to send data to multiple peer Destinations based on rules. Rules are evaluated in order, top->down, with the first match being the winner.

13.4. Pipelines

Data matched by a given Route is delivered to a Pipeline. Pipelines are the heart of Cribl Edge processing. Each Pipeline is a list of Functions that work on the data.

As with Routes, the order in which the Functions are listed matters. A Pipeline's Functions are evaluated in order, top->down.

Accessing Pipelines

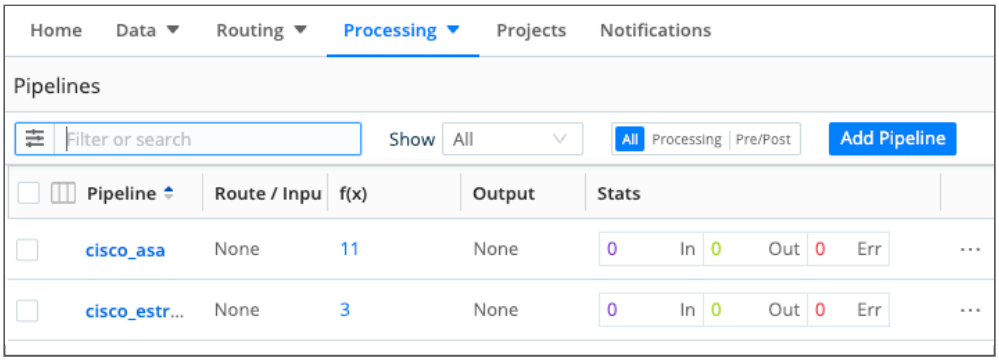
In single-instance deployments, select **More** > **Pipelines**. In distributed deployments, select **Manage** from the top nav, then select a **Fleet** to configure. From a **Fleet's** top nav, select **More** > **Pipelines**.

After you've clicked into a Pipeline, the right Preview pane adds a Pipeline Status tab, which you can click to see the Pipeline's events throughput.

Adding Pipelines

To create a new Pipeline, or to import an existing Pipeline to a different Cribl Edge instance, click **Add Pipeline** at the upper right. The resulting menu offers three options:

- **Create Pipeline**: Configure a new Pipeline from scratch, by adding Functions in Cribl Edge's graphical UI.
- **Import from File**: Import an existing Pipeline from a `.json` file on your local filesystem.
- **Import from URL**: Import an existing Pipeline from `.json` file at a remote URL. (This must be a public URL ending in `.json` – the import option doesn't pass credentials to private URLs – and the target file must be formatted as a valid Pipeline configuration.)



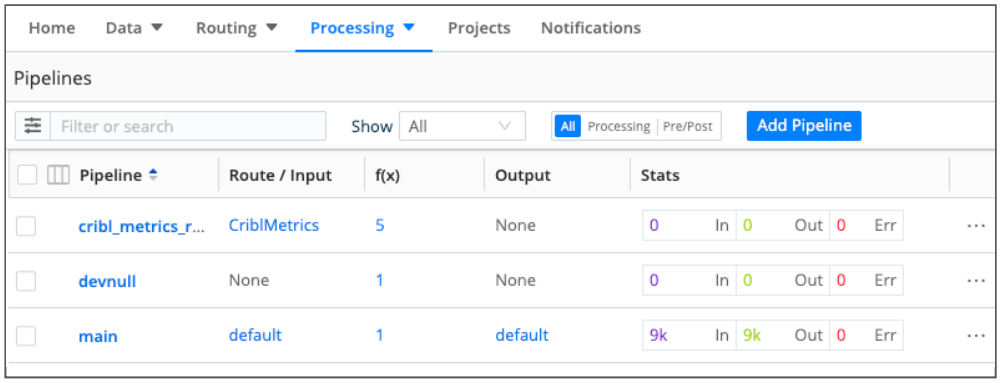
Creating or importing a Pipeline

To export a Pipeline, see Advanced Mode (JSON Editor).

To import or export a Pipeline along with broader infrastructure (like Knowledge Objects and/or sample data files), see Packs.

How Do Pipelines Work

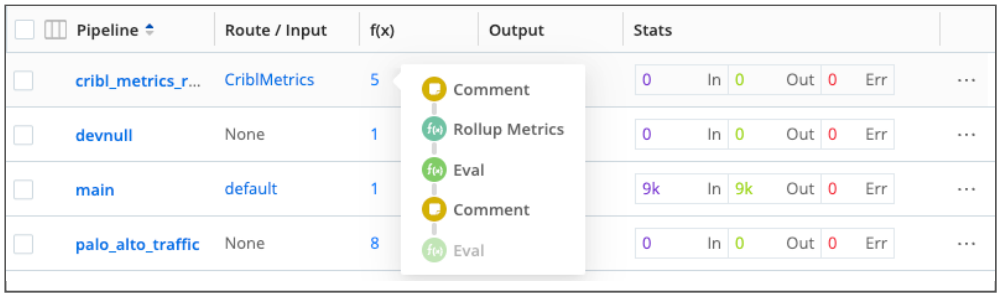
Events are always delivered to the beginning of a Pipeline via a Route. The data in the **Stats** column shown below are for the last 15 minutes.



Pipelines and Route inputs

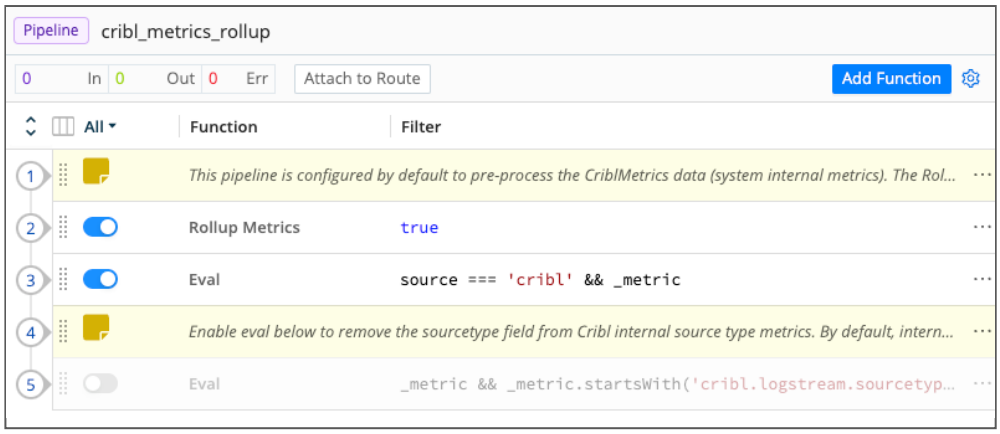
You can press the] (right-bracket) shortcut key to toggle between the Preview pane and an expanded Pipelines display. (This shortcut works when no field has focus.)

In the condensed Pipelines display above, you can also hover over any Pipeline's **Functions** column to see a horizontal preview of the stack of Functions contained in the Pipeline:



Preview on hovering over the top Pipeline (highlighted in gray)

Within the Pipeline, events are processed by each Function, in order. A Pipeline will always move events in the direction that points outside of the system. This is on purpose, to keep the design simple and avoid potential loops.



Pipeline Functions

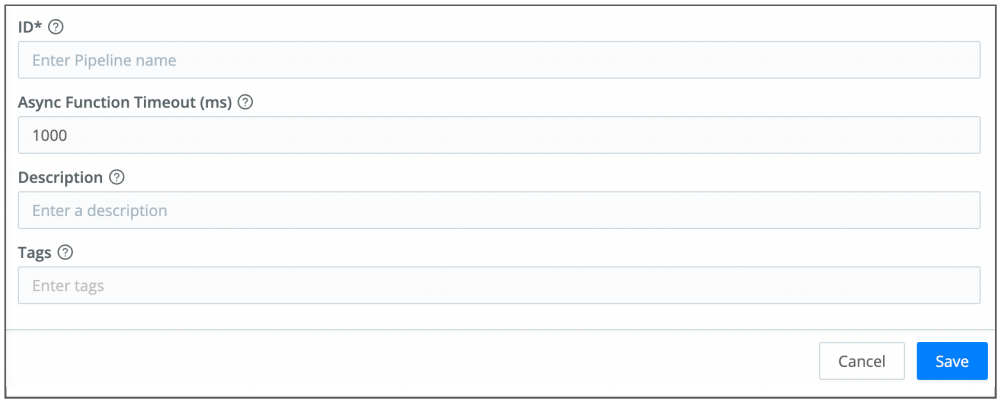
Use the **Attach to Route** link at upper left to associate a new Pipeline with a Route.

You can streamline a complex Pipeline's display by organizing related Functions into Function groups.

Pipeline Settings

Click the gear button at the top right to open the Pipeline's Settings. Here, you can:

- Use the **Async function timeout (ms)** to set the maximum amount of processing time, in milliseconds, that a Function is allowed to take before it is terminated. This prevents a Function from causing undesirable delays in your Pipeline (for example, a Lookup Function taking too long to process a large lookup file).
- Use the **Tags** field to attach arbitrary labels to the Pipeline. Once attached, you can use these tags to filter/search and group Pipelines.



Pipeline Settings

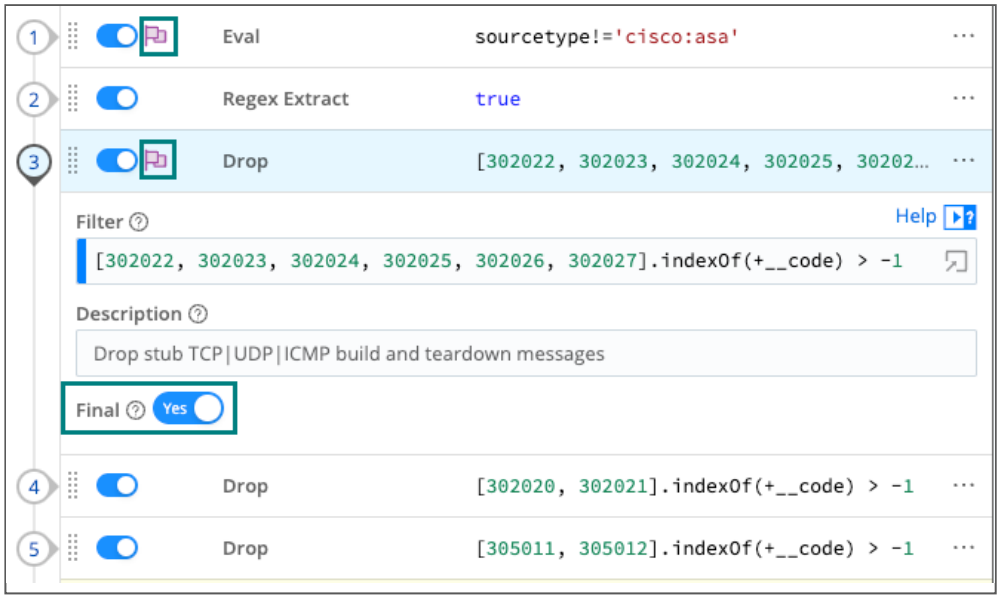
The `Final` Toggle

The `Final` toggle in Function settings controls what happens to the results of a Function.

When `Final` is toggled to `No` (default), events will be processed by this Function, **and** then passed on to the next Function below.

When `Final` is toggled to `Yes`, the Function "consumes" the results, meaning they will not pass down to any other Function below.

A flag in Pipeline view indicates that a Function is `Final`.



The Eval and first Drop Functions bearing the Final flag

Advanced Mode (JSON Editor)

Once you've clicked the gear button to enter Pipeline Settings, you can click **Edit as JSON** at the upper right to edit the Pipeline's definition in a JSON text editor. In this mode's editor, you can directly edit multiple values. You can also use the **Import** and **Export** buttons here to copy and modify existing Pipeline configurations, as `.json` files.

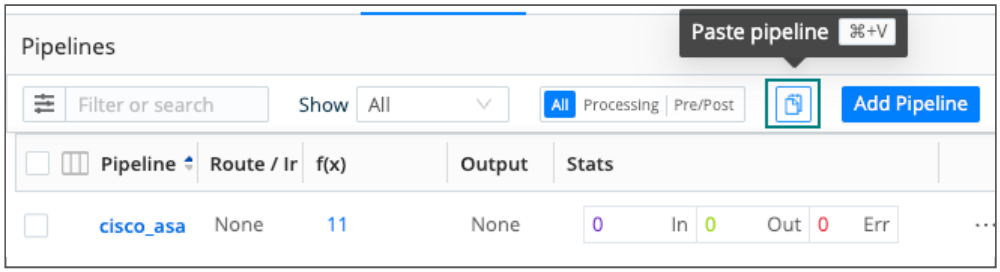
Advanced Pipeline Editing

Click **Edit in GUI** at upper right to return to the graphical Pipeline Settings page; then click **Back to <pipeline-name>** to restore the graphical Pipeline editor.

Pipeline Actions

Click a Pipeline's Actions (**...**) menu to display options for copying or deleting the Pipeline.

Copying a Pipeline displays the confirmation message and the (highlighted) Paste button shown below.



Paste button for copied Pipeline

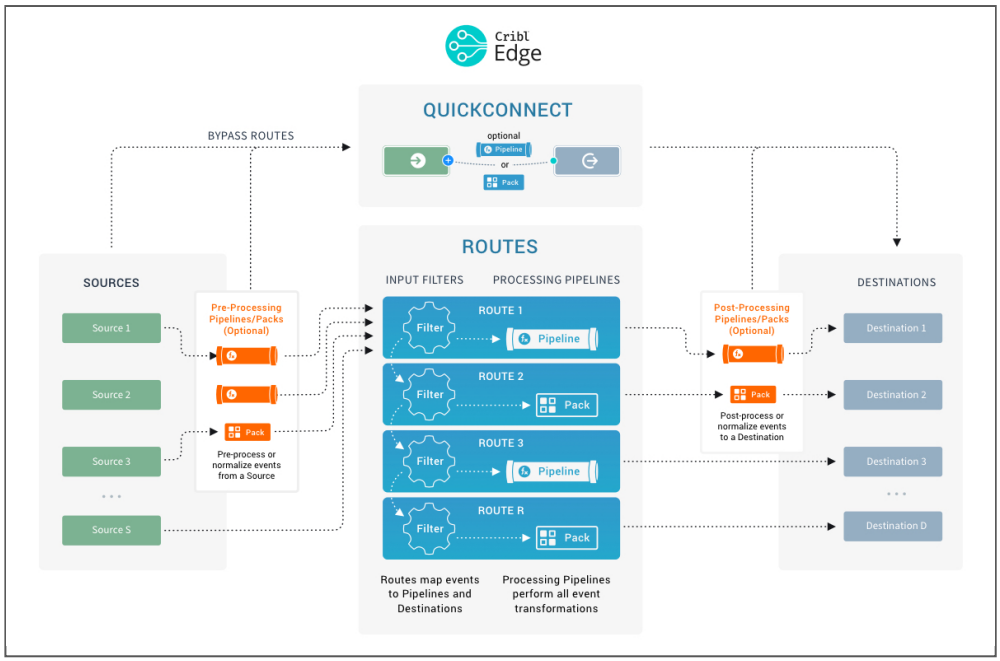
Pasting prompts you to confirm, or change, a modified name for the new Pipeline. The result will be an exact duplicate of the original Pipeline in all but name.

Chaining Pipelines

You can use the Chain Function to send the output of a Pipeline to another Pipeline or Pack. There are scope restrictions within Packs, and general guardrails against circular references.

Types of Pipelines

You can apply various Pipeline types at different stages of data flow. All Pipelines have the same basic internal structure (a series of Functions) – the types below differ only in their position in the system.



Pre-processing, processing, and post-processing Pipelines

Pre-Processing Pipelines

These are Pipelines that are attached to a Source to condition (normalize) the events **before** they're delivered to a processing Pipeline. They're optional.

Typical use cases are event formatting, or applying Functions to **all** events of an input. (E.g., to extract a `message` field before pushing events to various processing Pipelines.)

You configure these Pipelines just like any other Pipeline, by selecting **Pipelines** from the top menu. You then attach your configured Pipeline to individual Sources, using the Source's **Pre-Processing** > **Pipeline** drop-down.

Fields extracted using pre-processing Pipelines are made available to Routes.

Processing Pipelines

These are "normal" event processing Pipelines, attached directly to Routes.

Post-Processing Pipelines

These Pipelines are attached to a Destination to normalize the events before they're sent out. A post-processing Pipeline's Functions apply to **all** events exiting to the attached Destination.

Typical use cases are applying Functions that transform or shape events per receiver requirements. (E.g., to ensure that a `_time` field exists for all events bound to a Splunk receiver.)

You configure these Pipelines as normal, by selecting **Pipelines** from the top menu. You then attach your configured Pipeline to individual Destinations, using the Destination's **Post-Processing** > **Pipeline** drop-down.

You can also use a Destination's **Post-Processing** options to add **System Fields** like `cribl_input`, identifying the Cribl Edge Source that processed the events.

Best Practices for Pipelines

Functions in a Pipeline are equipped with their own filters. Even though filters are not required, we recommend using them as often as possible.

As with Routes, the general goal is to minimize extra work that a Function will do. The fewer events a Function has to operate on, the better the overall performance.

For example, if a Pipeline has two Functions, **f1** and **f2**, and if **f1** operates on `source 'foo'` and **f2** operates on `source 'bar'`, it might make sense to apply `source=='foo'` versus `source=='bar'` filters on these two Functions, respectively.

13.5. Packs

Packs enable Cribl Edge administrators and developers to pack up and share complex configurations and workflows across multiple Worker Groups, or across organizations.

Packs = Portability

With a Cribl Edge deployment of any size, using Packs can simplify and accelerate your work. Packs can also accelerate internal troubleshooting, and accelerate working with Cribl Support, because they facilitate quickly replicating your Cribl Edge environment.

For example, where a Pipeline's configuration references Lookup file(s), Cribl Edge will import the Pipeline only if the Lookups are available in their configured locations. A Pack can consolidate this dependency, making the Pipeline portable across Cribl Edge instances. If you import the Lookups into the Pack, you can develop and test a configuration, and then port it from development to production instances, or readily deploy it to multiple Fleets.

We don't claim to have brokered world peace here, but we do modestly hope to promote a stable, prosperous Pax Criblatica for the Cribl Edge ecosystem.

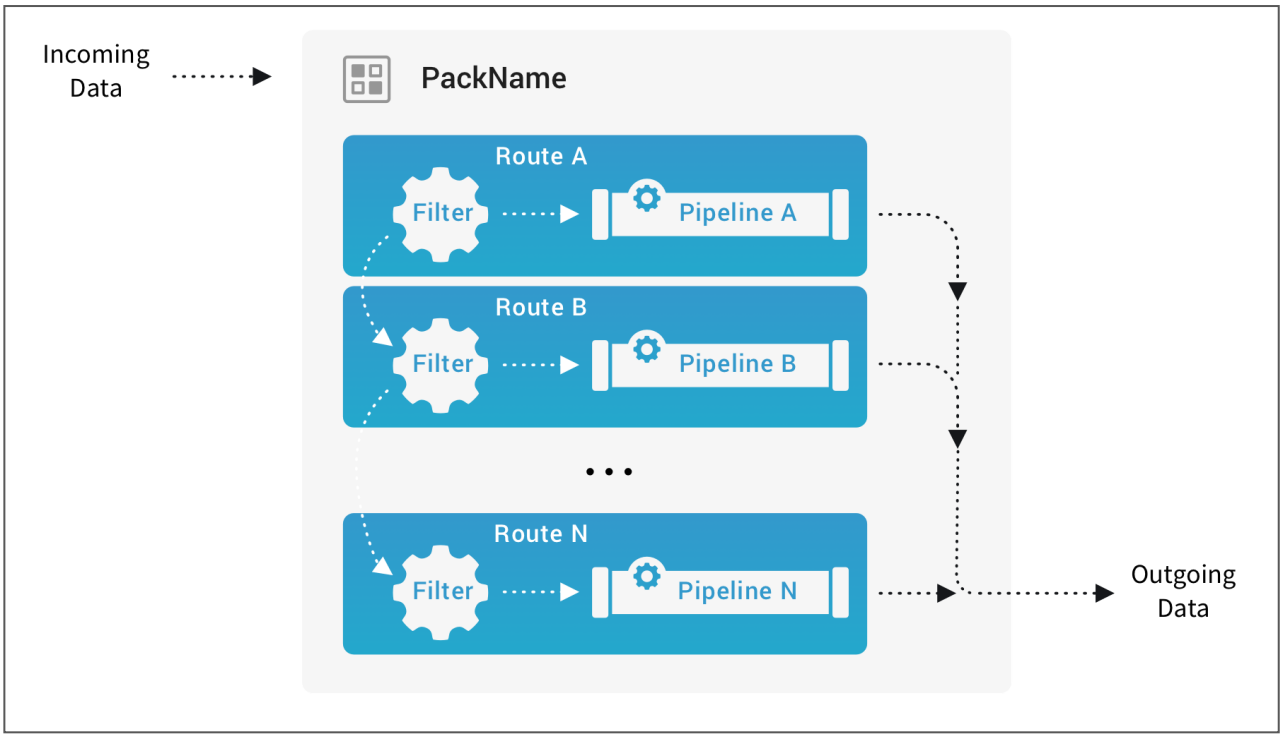
What Is a Pack?

Packs are implemented as a user interface (described on this page) and as a `.crbl` file format.

What's in a Pack?

Currently, a Pack can pack up everything between a Source and a Destination:

- Routes (Pack-level)
- Pipelines (Pack-level)
- Functions (built-in and custom)
- Sample data files
- Knowledge objects (Lookups, Parsers, Global Variables, Grok Patterns, and Schemas)



A Pack with internal Routes & Pipelines; no Knowledge or samples

As the above list suggests, a Pack can encapsulate a whole set of infrastructure for a given use case.

Packs don't have access to Knowledge objects (Lookups, Global Variables, etc.) that are defined outside the Pack. Knowledge objects defined **within** a Pack are not available to Cribl Edge resources outside the Pack. In either scenario, you must explicitly duplicate the Knowledge objects in the opposite scope.

What's Not in a Pack?

Sources, Collectors, and Destinations are external to Packs, so you can't specify them within a Pack. This excludes a few other things:

- Routes configured within a Pack can't specify a Destination.
- Packs can't include Event Breakers, which are associated with Sources. (However, you can instead use the Event Breaker Function in Packs' contained Pipelines.)

You connect a Pack with a Source and Destination by attaching it to a Route (see below), just as you'd attach a Pipeline.

Where Can I Get Some Packs?

Easy now. See The Cribl Packs Dispensary™ below.

Using Packs

These instructions cover using predefined Packs, as well as creating and modifying Pack configurations.

Version Compatibility

Packs created or modified in Cribl Edge 4.0.x cannot be used in any pre-4.0 version of Cribl Edge. (If you try, you'll see a `should NOT have additional properties` error.) To avoid this problem, Cribl recommends that you upgrade to Cribl Edge 4.0.x or later.

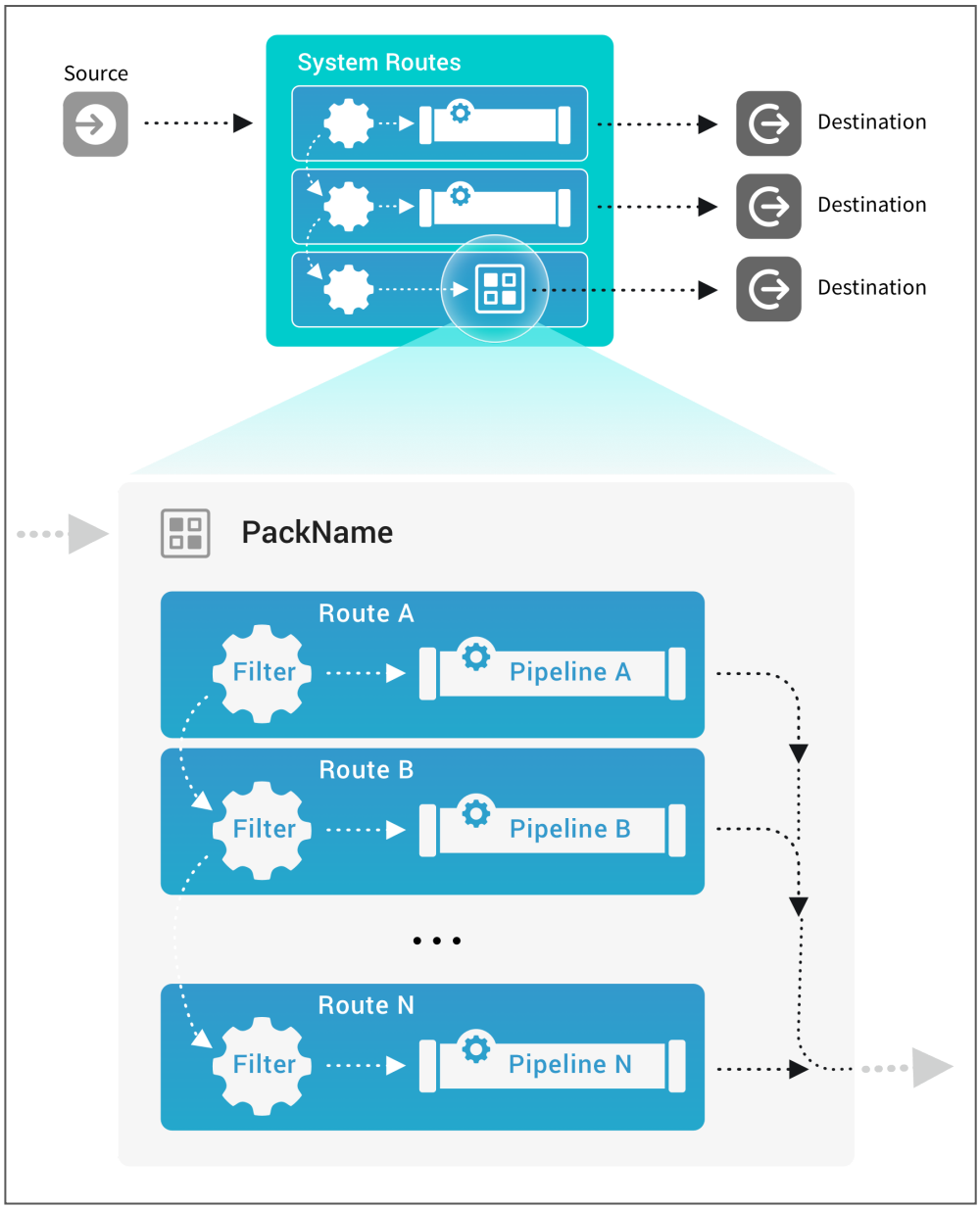
For compatibility questions about any individual Packs, please contact us in Cribl Community Slack's `#packs` channel.

Where Can I Use Packs?

Wherever you can reference a Pipeline, you can specify a Pack:

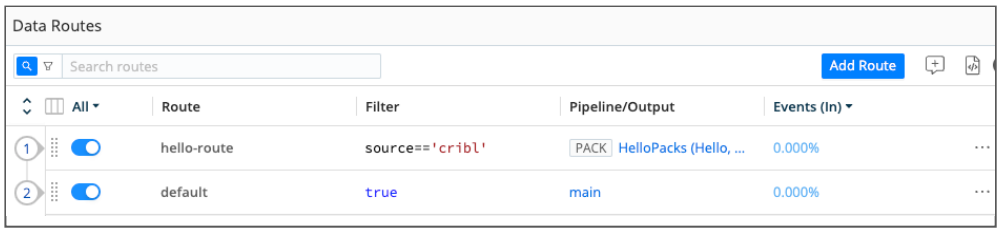
- In Sources, where you attach pre-processing Pipelines.
- In Destinations, where you attach post-processing Pipelines.
- In Routes, in the Routing table's **Pipeline/Output** column.

This expanded view shows how a Pack can replace a Pipeline in a Route:



A Pack snaps into Cribl Edge like an enhanced Pipeline

Packs are distinguished in the display with a **PACK** badge, as you can see here in the Routing table:



PACKs badged in Routing table's Pipeline column

The **PACK** badge is also displayed when you click into a resource – shown here on one of the Routes from the above table:

PACK badge on a Pack connected to a Route

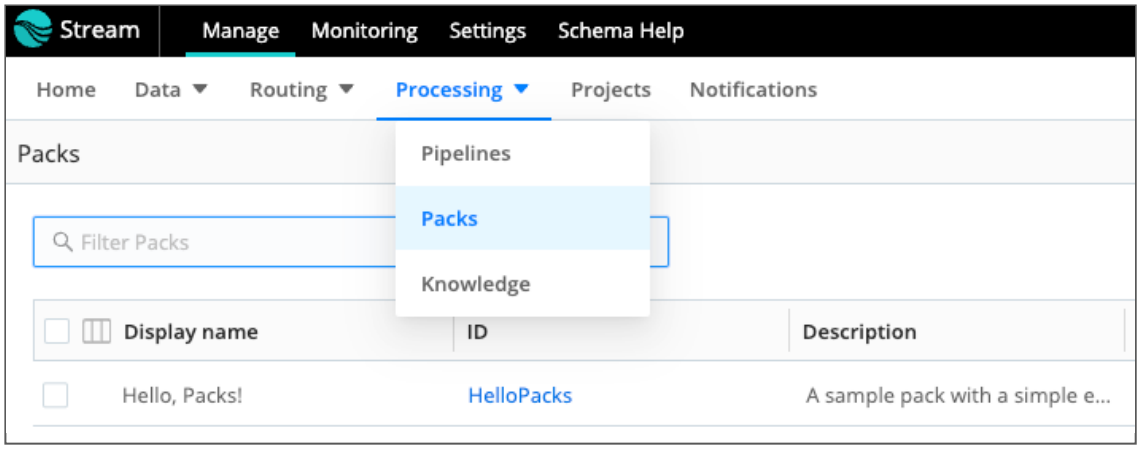
Cribl Edge's **Monitoring** page includes a **Packs** link where you can monitor Packs' throughput.

Accessing Packs

You access Packs differently, depending on your deployment type.

Single-Instance Deployment

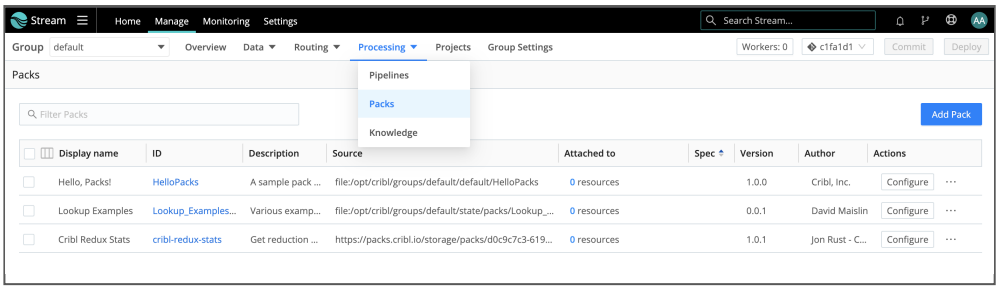
In a single-instance deployment, Packs are global. From Cribl Edge's top-level navigation, just select **Processing** > **Packs**. On the filesystem, Packs (including those that you add) are stored at `$CRIBL_HOME/default/`.



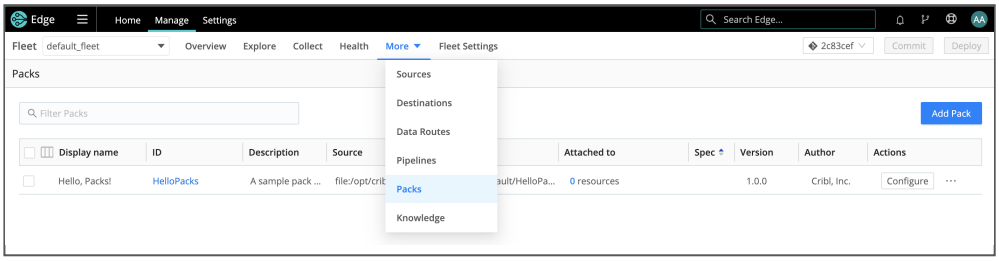
Packs, single-instance navigation

Distributed Deployment

In a distributed deployment, Packs are associated with (and installed within) Fleets. Select **Manage**, and then click into the Fleet you want to manage (for example, `default`). Next, from that Fleet's top nav, select **Processing** > **Packs** (Stream) or **More** > **Packs** (Edge).



Stream Group > Manage Packs page



Edge Fleet > Manage Packs page

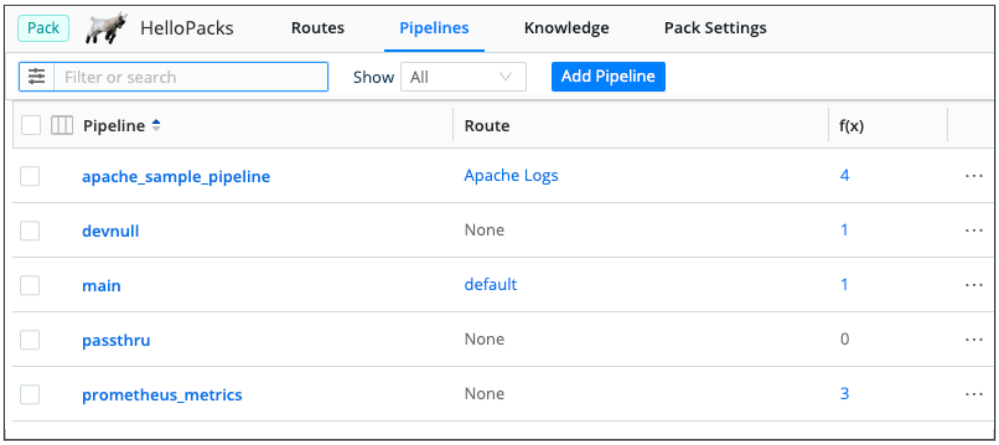
Each Group's Packs are stored at at `$CRIBL_HOME/groups/<group-name>/default/`. (In a typical installation, you'll find the starter `HelloPacks` Pack at `/opt/cribl/groups/default/default/`.)

By design, you can readily share Packs **across** Fleets by copying them between each other.

Getting Started with Packs

To unpack Packs, use the above instructions (per deployment type) to navigate to the **HelloPacks** example Pack shipped with Cribl Edge. On the **Manage Packs** page, click this Pack's row to see its configuration.

Click **Pipelines** on the Pack's submenu, and you'll see that the Pack includes `devnull`, `main`, and `passthru` Pipelines, corresponding to the default Pipelines provided at Cribl Edge's global level. This Pack also includes an Apache-specific sample Pipeline – click it to unpack that, too.

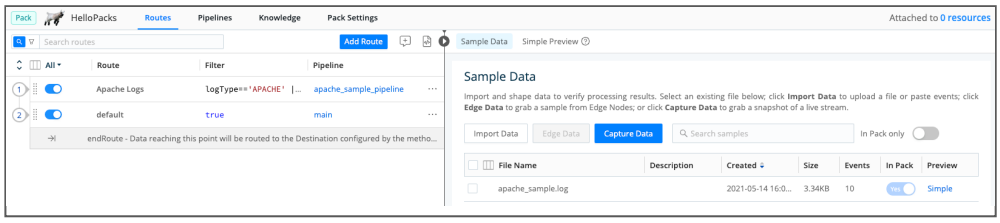


Pipelines within example Pack

Click **Routes** on the Pack's submenu, and you'll see that this Pack also provides both a `default` and an Apache-specific Route.

Configuring a Pack

Once loaded, each Pack displays a submenu with familiar links: **Routes**, **Pipelines**, **Knowledge**, and **Settings** on the left pane, along with **Sample Data**, and **Preview Simple** on the right. The Pack's submenu is a subset of Cribl Edge's top nav.



Configuring a Pack

The left pane's submenu links give you access to configuration objects specific to this Pack.

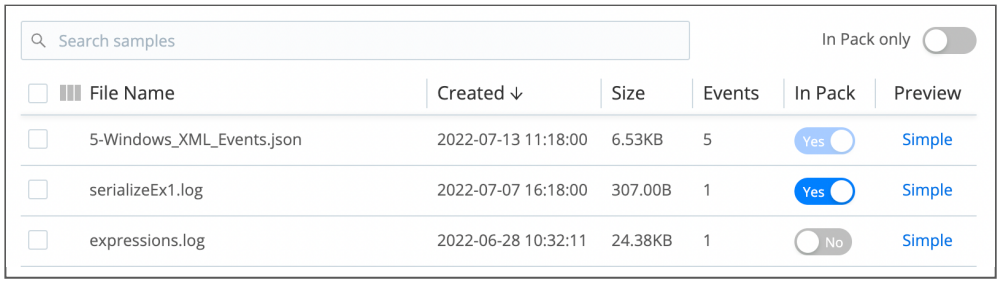
Sample Data

The right pane defaults to displaying all sample data files available on your Cribl Edge instance. If you prefer to filter only sample files internal to the Pack, toggle **In Pack only** to the right.

If you add sample data files via this Pack UI, they will be internal to that Pack. Each sample file here displays its own **In Pack** toggle on its row, which works as follows:

A light-blue toggle is locked, meaning that this sample file is internal to the Pack. It will export with the Pack. If you want to make this sample available across Cribl Edge, you'll need to also add it via the global right preview pane (accessed from **Routing** > **Data Routes** or **Processing** > **Pipelines**).

A grayed-out or dark-blue toggle means that this sample file is global to Cribl Edge. It is available to this Pack. Toggle this to `Yes` (dark blue) if you want the sample file to export along with the Pack.



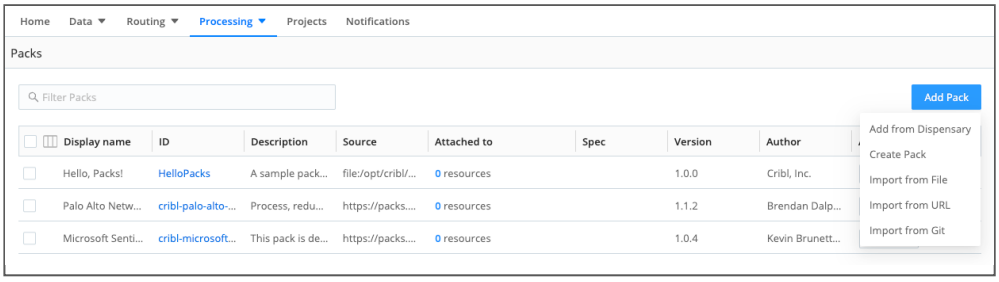
Sample file in Pack

Basically, you can manipulate all the options here as you'd work with their big sister or brother in Cribl Edge's global navigation.

Importing or Upgrading a Pack

To import a new Pack, or an updated version of an existing Pack, from your filesystem:

1. Navigate to the **Manage Packs** page.
2. Click **Add Pack** at the upper right.
3. Select your desired **Add/Import** source: Dispensary, File, URL, or Git repo.
4. Follow the above links to details on each of these options.



Importing a Pack

Custom Functions

Packs can include Pipelines containing custom functions, which can (in turn) run arbitrary JavaScript. Before you install a Pack, make sure it comes from a provider you trust, such as the Cribl Packs Dispensary or your own organization.

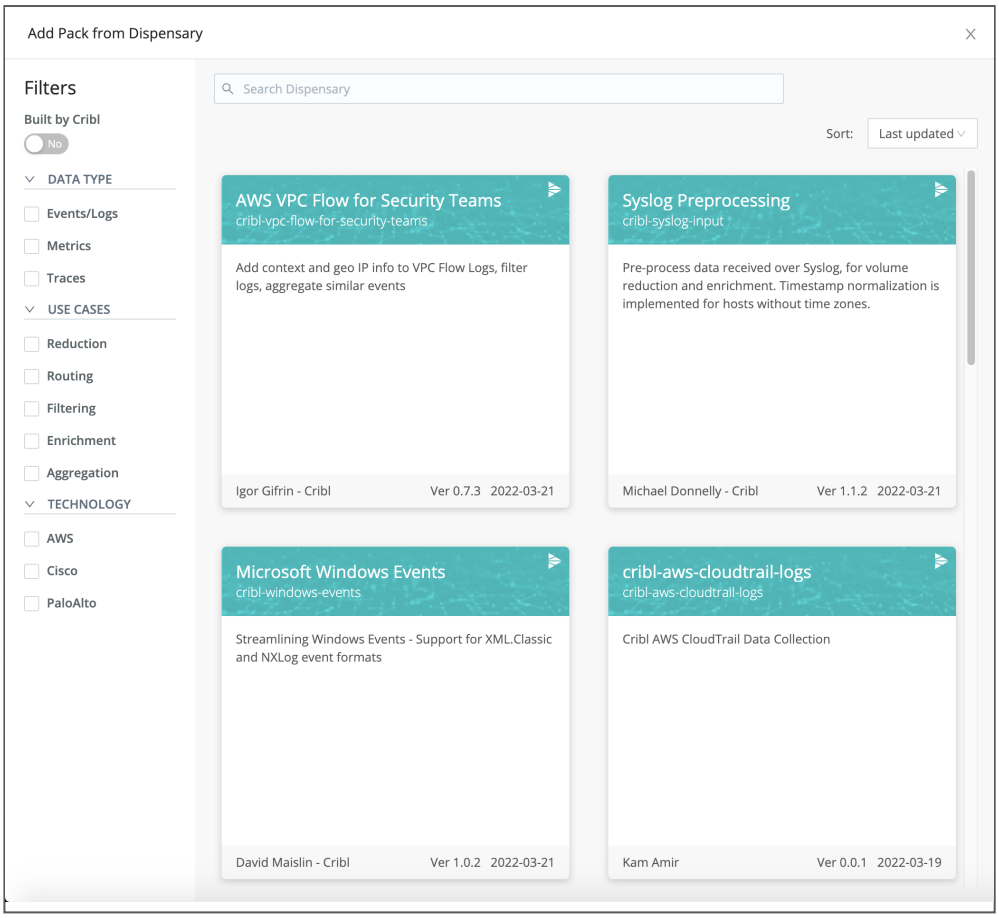
As an additional protection layer, all Pack import modals provide an **Allow custom functions** toggle. In the toggle's default `No` position, if Cribl Edge detects custom functions in the specified Pack, it will block the import with an error message. If you trust the Pack's provider,set the toggle to `Yes`, and the import will proceed normally.

The Cribl Packs Dispensary™

You might be wondering, "Where can I find a reliable source of Packs that add useful features to Cribl Edge, vetted for safety?"

Well, Cribl is proud to point you to the Cribl Packs Dispensary™. Here, Cribl's own solutions engineers have seeded several strains of high-productivity Cribl Edge configurations. Because the Packs Dispensary™ is a place to share good stuff, we expect many new hybrids to sprout from the community. Cribl will test and curate submissions to ensure the quality of the repo's contents.

You can install Dispensary Packs directly through Cribl Edge's UI, as outlined in Add from Packs Dispensary below.



Cribl Packs Dispensary™ (as displayed in Cribl Edge's **Add** drawer)

Interested in publishing your own Packs on the Cribl Packs Dispensary™? See Publishing a Pack.

Add from Packs Dispensary

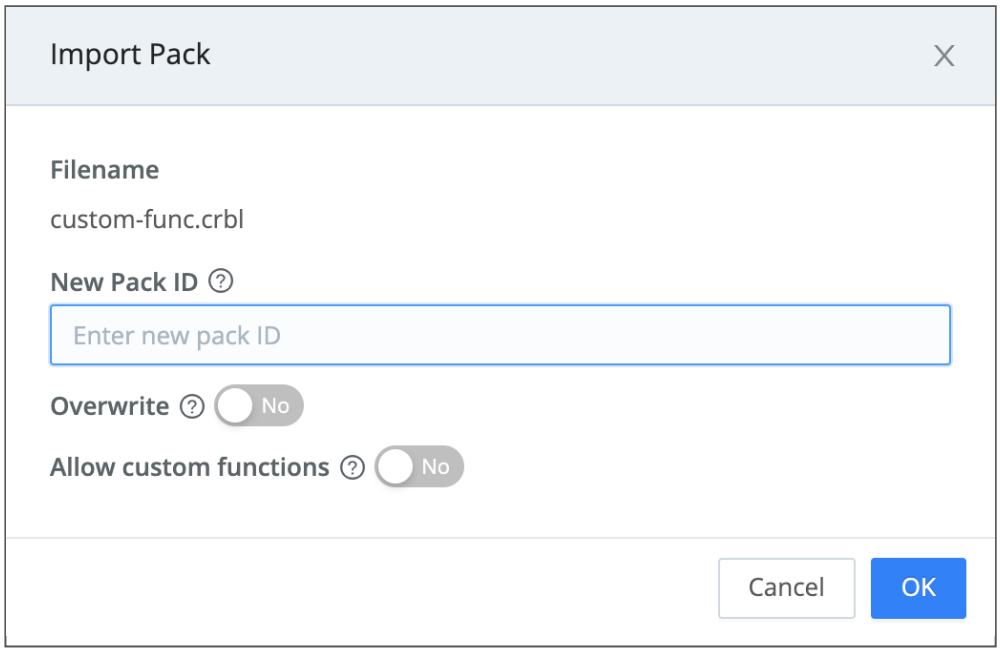
To add a Pack from the Cribl Packs Dispensary™ sharing site:

1. From the **Manage Packs** page's **New Pack** submenu, select **Add from Dispensary**.
2. The Packs Dispensary will open in a drawer, as shown in the screenshot above.
3. Using the drawer's controls, browse or search for the Pack(s) you want. (You can use the check boxes at the left to filter by data type, use case, and technology.)
4. Click any Pack's tile to display its details page with its README. This will typically outline the Pack's purpose, compatibility, requirements, and installation.
5. To proceed, click **Add Pack** on this page.
6. That's it! You'll see a banner confirming that the Pack is now installed.

Import from File

To import a Pack (`.crbl` file) from your local filesystem:

1. From the **New Pack** submenu, select **Import from File**.
2. From the resulting File Open dialog, select the file to import.
3. Optionally, give the pack an explicit, unique **New Pack ID**. (For details about this option, see Upgrading an Existing Pack below.)
4. Where appropriate (see just above), enable **Allow custom functions**.
5. Click **OK** to confirm the import.

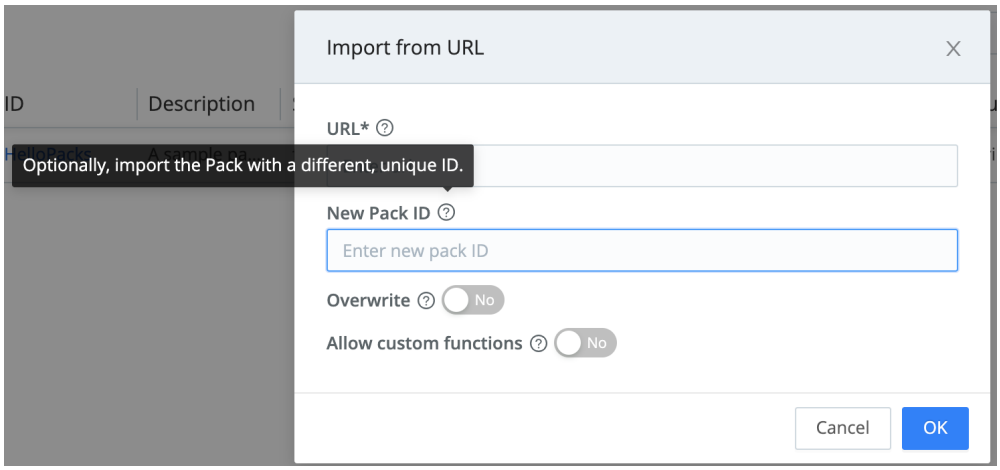


Importing from a file

Import from URL

To import a Pack from a known, public or internal, URL:

1. From the **New Pack** submenu, select **Import from URL**.
2. Enter a valid URL for the Pack's source. (This field's input is validated for URL format, but not for accuracy, before you submit the modal.)
3. Optionally, give the pack an explicit, unique **New Pack ID**. (See Upgrading an Existing Pack.)
4. Where appropriate, enable **Allow custom functions**. (See Custom Functions.)
5. Click **OK** to confirm the import.



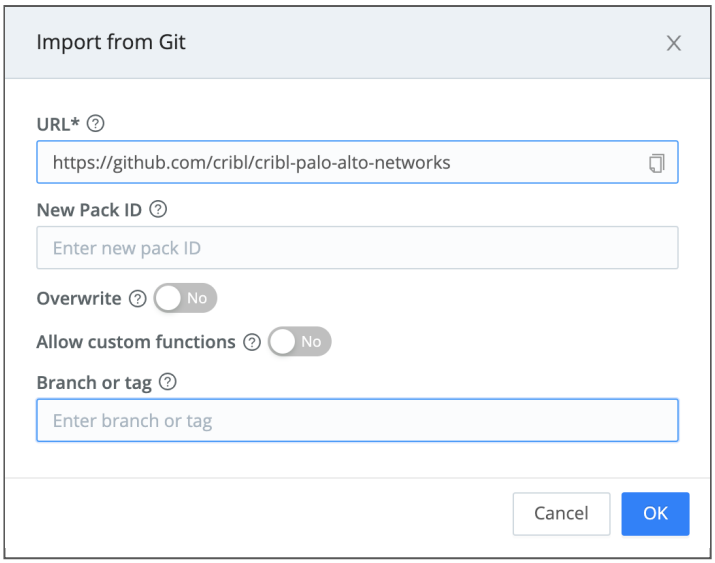
Confirming file import from URL

To import a Pack from a public URL, Cribl Edge's Leader Node (or single instance) requires Internet access. A distributed deployment's Leader can then deploy the Pack to Workers even if the Workers lack Internet access.

Import from Git Repos

To import a Pack from a known public or private Git repo:

1. From the **New Pack** submenu, select **Import from Git**.
2. Enter the source repo's valid URL.
   This field's input is validated for URL format, but not for completeness or accuracy, before you submit the modal. When targeting a private repo, use the format: `https://<username>:<token/password>:<repo-address>`. Public repos need only `https://<repo-address>`, as shown in the example below.
3. Optionally, give the pack an explicit, unique **New Pack ID**. (See Upgrading an Existing Pack.)
4. Optionally, enter a **Branch or tag** to filter the import source using the repo's metadata. You can specify a branch (such as `master`) or a tag (such as a release number: `0.5.1`, etc.).
5. Where appropriate (see Custom Functions), enable **Allow custom functions**.
6. Click **OK** to confirm the import.



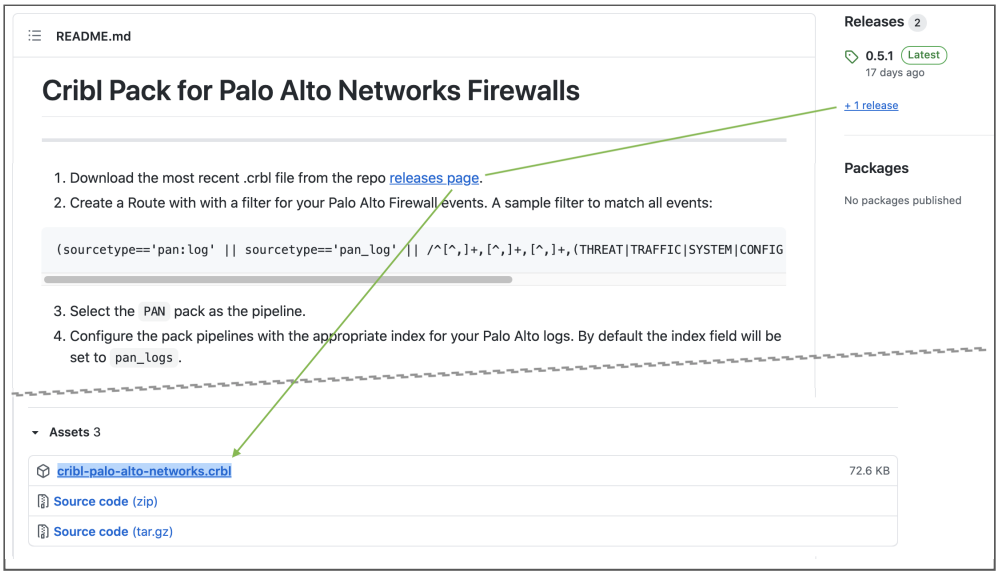
Importing from a Git repo

To import a Pack from a public repo, Cribl Edge's Leader Node (or single instance) requires Internet access. A distributed deployment's Leader can then deploy the Pack to Workers even if the Workers lack Internet access.

Dispensary GitHub Repo

One authoritative public repo is the Cribl Pack Dispensary on GitHub. (This is the precursor to the Cribl-hosted Cribl Packs Dispensary™ site.)

You can install Dispensary Packs directly through Cribl Edge's UI, as outlined in Import from Git Repos above. However, if you prefer, you can click through to any Dispensary repo's release page, download the corresponding `.crbl` file, and then upload the file into Cribl Edge.



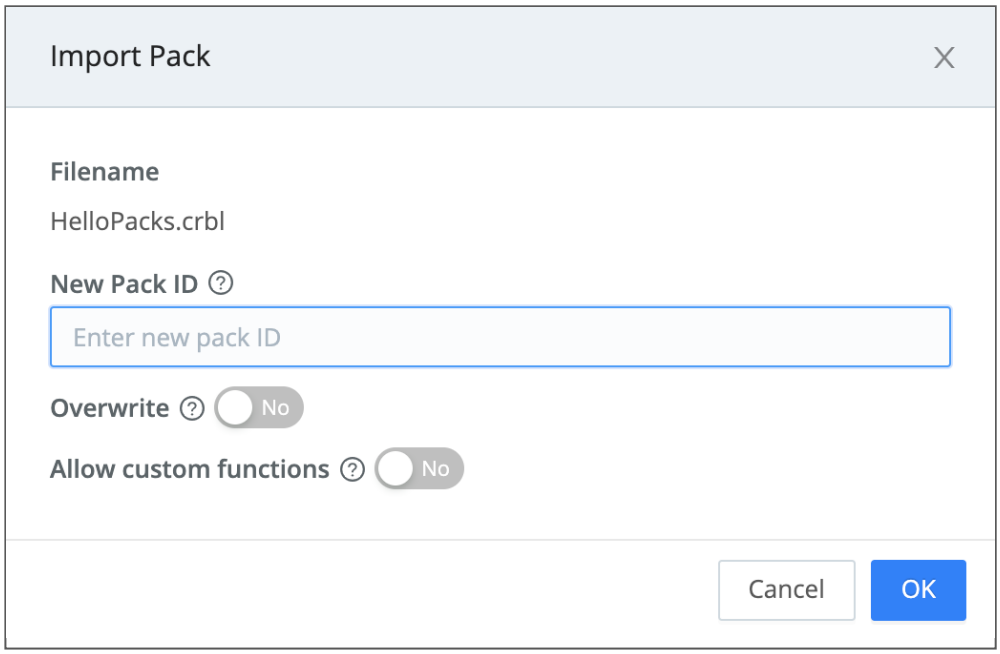
Downloading a `.crbl` file from the Cribl Pack Dispensary's Web UI

If you've posted completed Packs to our GitHub repo, we encourage you to now submit them to our new Cribl Packs Dispensary™ site. See Publishing a Pack.

Upgrading an Existing Pack

Each Pack that is installed within a given Fleet (or single-instance deployment) must have a unique ID. The ID is based on the Pack's internal configuration – not its container's file name, nor on its Display name.

If you import a Pack whose internal ID matches an installed Pack – whether an update, or just a duplicate – you'll be prompted to assign a unique **New Pack ID** to import it as a separate Pack.



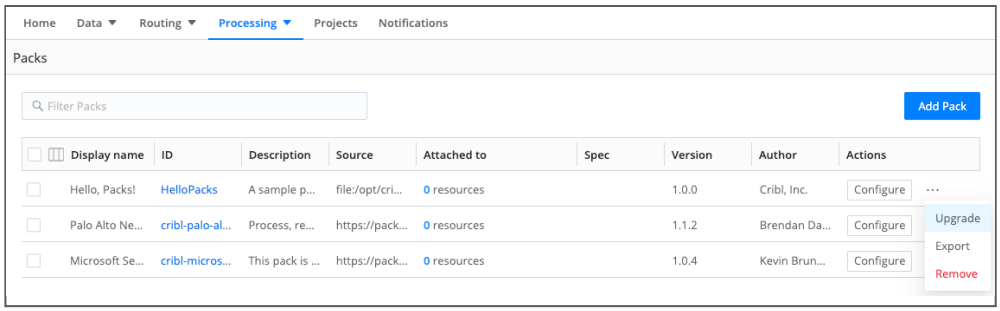
Renaming a Pack on import

You'll also have the option to **Overwrite** the installed Pack, reusing the same ID.

If you toggle this option to `Yes`, the imported Pack will completely overwrite your existing Pack's configuration.

Each Pack within a Cribl Edge instance must have a unique **Pack ID**, so you cannot share an ID between two (or more) installed Packs.

To explicitly upgrade an existing Pack, you can instead click the menu icon on its row and select **Upgrade**.



Upgrading an existing Pack

If you've modified an installed Pack, Cribl Edge will block the overwrite of the Pack, to prevent deletion of your locally created resources.

When upgrading a Pack, Cribl recommends:

- Import the updated Pack under a new name that includes the version number (e.g., `cribl-syslog-input-120`). This allows you to review and adjust new functionality against currently–deployed configurations.
- Do a side–by–side comparison of the previous and new versions of the Pack – remember to review all comments in the new Pack. Doing this side-by-side comparison allows you to copy Function expressions and other settings from the current version into the same fields in the new version.
- Enable or disable any Functions in the new Pack as necessary.
- Update any Routes, Pipelines, Sources, or Destinations that use the previous Pack version to reference the new Pack.
- If the Pack includes any user–modified Knowledge objects (e.g., Lookup files), be sure to copy the modified files locally for safe keeping before upgrading the Pack. After installing the upgrade, copy those files over the Pack upgrade's default versions.
- Test, test, and test!
- Commit and Deploy.

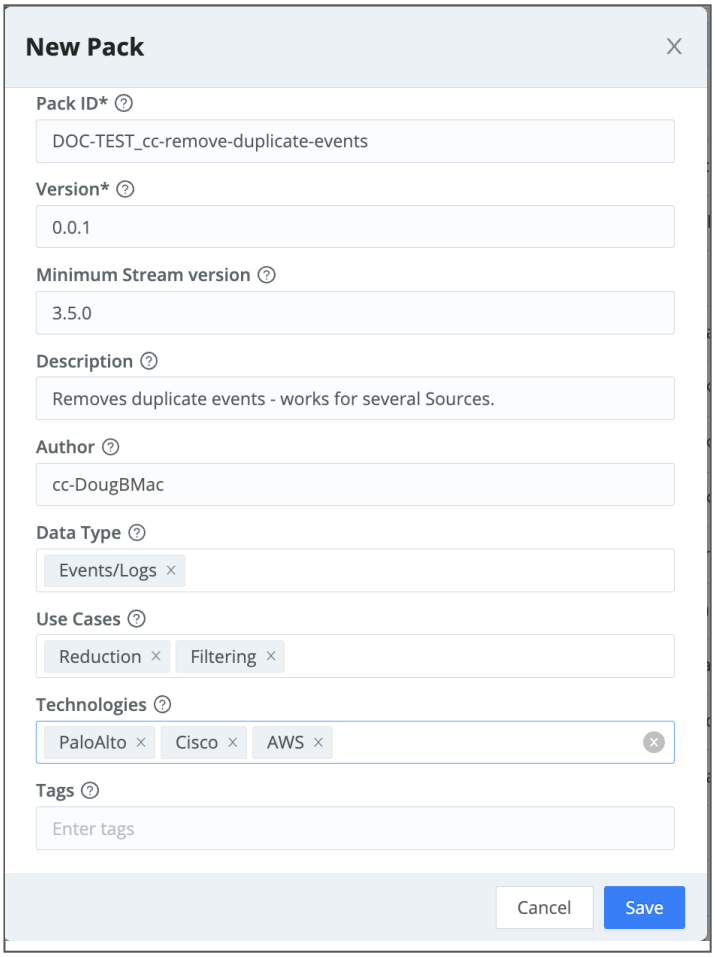
Creating a Pack

You can create a new Pack from scratch, to consolidate and export multiple Cribl Edge configuration objects:

1. Navigate to the **Manage Packs** page.
2. Click **Add Pack**.
3. From the submenu, select **Create Pack**.
4. In the resulting **New Pack** modal, fill in a unique **Pack ID** and other details.

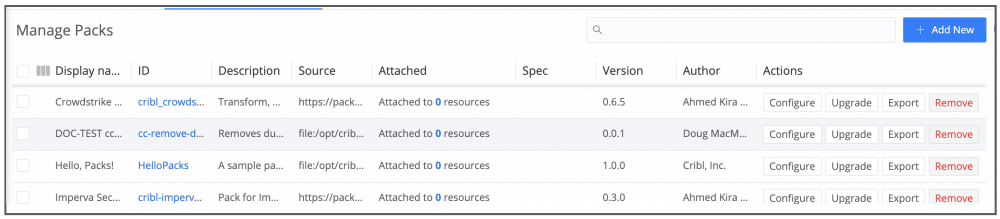
- Each Pack within a Cribl Edge instance must have a separate **Pack ID**, but you can assign arbitrary **Display name**s.
- **Version** is a required field identifying the Pack's own versioning.
- **Minimum Stream version** is an optional field specifying the lowest compatible version of Cribl Edge software.
- **Description** and **Author** are optional identifiers.
- **Data type**, **Use cases**, and **Technologies** are optional combo boxes. You can insert one or multiple keywords to help users filter Packs that you post publicly on the Cribl Packs Dispensary™.
- **Tags** are optional, arbitrary labels that you can use to filter/search and organize Packs.

5. Click **Save**.



Creating a Pack

6. On the **Manage Packs** page, click the new Pack's row to open the Pack.



Manage Packs page

7. Use the standard Cribl Edge controls (see above) to configure and save the infrastructure you want to pack up. As you save changes in the UI, they're saved to the Pack.

If you'd like to share your Pack with the community of Cribl users, you can publish it on the Cribl Packs Dispensary™.

The Cribl Packs Dispensary™ site is designed for sharing completed Packs. If you want to collaborate with others on iteratively developing a Pack, Cribl recommends relying on our Dispensary GitHub Repo for the development phase.

If you have a Cribl.Cloud account, you can also collaborate there by inviting team members to your Cribl.Cloud Organization. See Inviting and Managing Other Users.

Once your Pack is ready to share, we encourage you to submit it to the Cribl Packs Dispensary™ site. If you already have completed Packs on our GitHub repo, bring them over here!

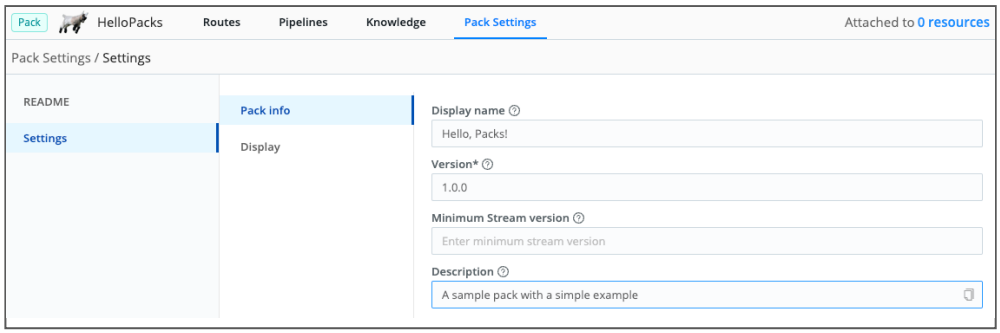
Modifying Pack Settings

You can update a Pack's metadata (Version, Description, Author, etc.) and display settings. If you're developing a new Pack to share, you'll want to use this interface to populate the Pack's README and display logo.

1. From the Pack's submenu, select **Pack Settings**. The left **README** tab will gain focus.
2. To populate the Pack's README file, toggle **View** to **Edit**, replace the placeholder markdown content, and **Save**.

Editing Pack's README

3. To update other metadata, click the left **Settings** tab.



Editing Pack's metadata

4. To add a Pack logo, click the Pack's **Settings** > **Display** left tab.
   Cribl recommends adding a logo to each custom Pack, to visually distinguish the Pack's UI from the surrounding Cribl Edge UI (as well as from other Packs). You can upload a `.png` or `.jpg`/`.jpeg` file, up to a maximum size of 2MB and 350x350px. Cribl recommends a transparent image, sized approximately 280x50px.

Copying a Pack

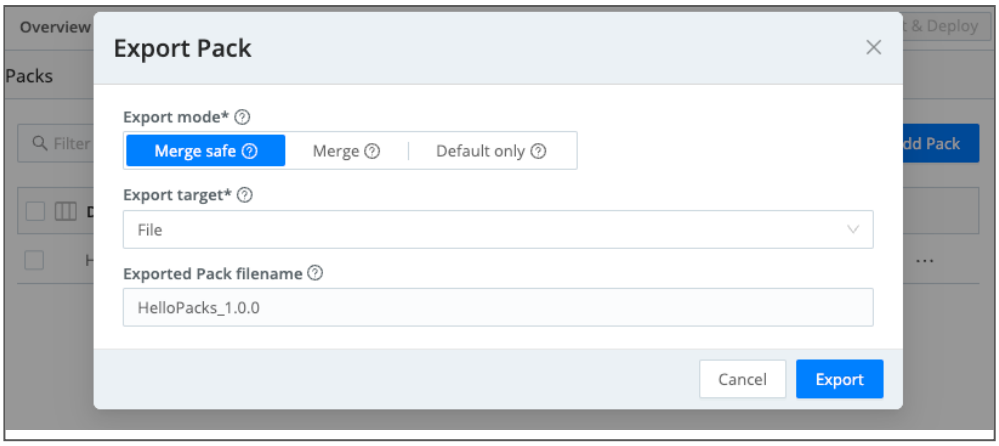
You can easily share Packs in a distributed deployment by copying them between Fleets.

1. You can copy one or more Packs.
   - To copy a single Pack: Select its menu icon on the Packs page and select **Copy to another Fleet**.
   - To copy multiple Packs in one operation: Select their check boxes and then select **Copy selected Packs to another Fleet**.
2. Select the Fleets to copy the Pack to.
3. Optionally, toggle **Overwrite** on to replace existing Packs with the same **Pack ID**.
4. Select **Copy**.

A status modal will provide a successful message or list any Packs that failed to copy.

Exporting a Pack

To export a newly created or modified Pack, click its menu icon on the Packs page and select **Export**.



Exporting a Pack

The resulting Export Pack modal provides the following options.

Export Mode

Select one of these three buttons:

- **Merge safe**: Attempt to safely merge local modifications into the Pack's default layer (original configuration), then export.
- **Merge**: Force-merge local modifications into the Pack's original configuration, then export.
- **Default only**: Export only the Pack's original configuration, without local modifications.

The **Merge safe** option is conservative, and will block the export where Cribl Edge can't readily merge conflicting, modified contents with the Pack's original contents:

Merge safe error

If you encounter an error like the example shown above, use the **Merge** or **Default only** export mode instead.

Exported Pack ID

Defaults to the Pack's current ID, with the version number appended. This is an opportunity to change the Pack's ID if you're exporting it as you develop a new version.

Managing Packs via API

You can perform Pack operations by running Cribl Edge API calls on the command line. This is required if you plan to automate Pack operations, e.g., in a CI/CD pipeline.

In this section, we'll walk through one scenario where running API calls on the command line works well: exporting a Pack from one Fleet and installing it into another. The two Worker Groups/Fleets do **not** need to have the same Leader Node.

About the Following Examples

- The API calls here include Fleet names as path parameters.
- The `curl` commands assume that you have set the `$token` environment variable to match the value of a bearer token. Of course, this is just one option for authentication. See the Authentication topic for others; adapt the example commands to suit your chosen approach.

Export via API

Adapt and run this **Export pack** API call, using the export mode of your choice:

```
GET /api/v1/m/<worker_group_name>/packs/<pack_name>/export?mode=merge
```

Export Example

Let's export a Pack named `goat-herd` from the `default` Fleet, and use the `>` redirect to write the exported Pack to a file named `goat-herd.crbl`:

```
curl -X GET -H "Authorization: Bearer $token" 'https://stream:9000/api/v1/m/default
```

This request returns an octet-stream attachment which is downloaded as a `.crbl` file. And voilà, you have exported your Pack.

Install via API

Installing the exported Pack in a different Fleet is a two-step process: First upload, then actually install.

Install via API – Step 1

Adapt and run this **Upload pack** API call, referencing the exported Pack file:

```
PUT /api/v1/m/<new_worker_group>/packs?filename=<pack_name>.crbl
```

Install Example – Step 1

We'll use our target Fleet name (in this example, it's `group420`). Then we need to specify the exported Pack contents as a file payload, using the `--data-binary` option to upload the binary data without modification. The `@` prefix tells `curl` that `goat-herd.crbl` is the path to the file, not the data itself.

```
curl -X PUT -H "Authorization: Bearer $token" 'https://stream:9000/api/v1/m/group42
```

This request returns a JSON object of the following form:

```
{"source":"pack_name.random_id.crbl"}
```

Install via API – Step 2

Adapt and run this **Install pack** API call:

```
POST /api/v1/m/<new_worker_group>/packs
```

Meanwhile, remember that this API call will need a payload – the JSON object returned by the previous API call.

Install Example – Step 2

We'll use the `curl -d` option to specify the JSON object payload. We'll add a new element to the object, whose key is `id`, and whose value is the Pack's new name in the new Fleet.

Here, the `goat-herd` Pack is renamed as `billys_pack`. (If you do not wish to rename the Pack, just omit the `id` element – but keep the `source` element.)

```
curl -X POST -H "Authorization: Bearer $token" -H "Content-Type: application/json"
```

Copy via API

To bulk-copy Packs between Worker Groups/Fleets, adapt and run this **Clone pack** API call, referencing a source Fleet, destination Fleet(s)/Fleet(s), and Pack(s).

```
POST /api/v1/packs/__clone__

{
  "srcGroup": "copy_from_this_worker_group_id",
  "dstGroups": [
    "destination_group_1",
    "destination_group_2",
    ...
  ],
  "packs": [
    "pack_id_1",
    "pack_id_2",
    ...
  ]
}
```

Copy Example

For example, to copy the Palo Alto Networks and Cisco ASA Packs from the default to `dc1-logs` and `dc2-logs` Fleets:

```
curl -X POST -H "Authorization: Bearer $token" -H "Content-Type: application/json"
```

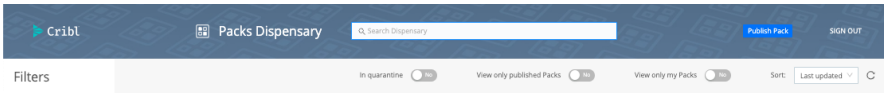
13.5.1. Packs Publication Standards

This page outlines the process for Cribl Community members to publish Cribl Edge Packs to the Cribl Packs Dispensary. It also lists standards that apply to all publicly available Community Packs.

Publication Overview

Publishing your Pack is a five-stage process:

1. **Prepare**
    - Prior to starting development, read the Pack Review Summary and Pack Review Checklist below. This will help you create a Pack that can easily pass the review process. Decide on the problems you're trying to solve, and determine the dataset(s), Sources, and Destinations you're going to work with.
2. **Develop your Pack**
    - Development is generally done in your own Cribl environment, using saved samples to build out pipelines. During this stage, you might consider working collaboratively within Cribl's Community Slack `#packs` channel. When the Pack is ready for review, export the `.crbl` file using the Cribl UI.
3. **Test your Pack**
    - Cribl strongly recommendeds that you test the Pack yourself, in your own environment. To test that all content was exported correctly, import your Pack into a fresh Cribl environment, such as a newly created Docker instance, or to a newly created Worker Group in your Cribl environment. Upload the Pack, then use the Pack Review Checklist to verify that the Pack will pass review.
4. **Submit your Pack for review**
    - Sign into, or create an account on, the Cribl Packs Dispensary site. You can use your Cribl.Cloud account's email address to create the Cribl Packs Dispensary account.
    - If you don't have a Cribl.Cloud account, Cribl automatically creates a new one for you when you create your Cribl Packs Dispensary account.
    - Once you sign in, you'll see the Dispensary screen with the **Publish Pack** button.



Packs Dispensary controls

    - Click **Publish Pack**, and then click **Upload Pack**.
    - Select the `.crbl` Pack file you want to publish, and click **Submit**. You will be prompted to agree to the terms of Cribl's Pack Developer Agreement. As a Pack author, you must electronically

acknowledge that you have read the Agreement, and that you intend to adhere to its requirements.

Once the Pack has been uploaded, Cribl will review your Pack. The review process typically takes a week or more. Once your Pack is reviewed, you'll receive an email from `packs@cribl.io` either informing you that it was accepted, or providing the reason(s) why it was rejected.

5. **Once your accepted Pack is published to the Cribl Packs Dispensary, it's time to celebrate!**

Pack Review Summary

To ensure consistency across published Packs, Cribl evaluates all submitted Packs against the checklist below. Any Pack that fails to satisfy the checklist's requirements will be rejected.

Cribl recommends:

- Check your Pack against the checklist **before** trying to submit the Pack. This way you can avoid having your Pack rejected for something easily fixed.
- Install the **Syslog Pre-processing** (`cribl-syslog-input`) Pack as a reference – it includes clear examples of each item on the checklist.

Pack Review Checklist

1. Naming

Every Pack must have both a Pack ID and a Display Name.

When you view Packs in the Dispensary, the Display Name appears above the Pack ID. For example, **Syslog Pre-processing** is a Display name while `cribl-syslog-input` is a Pack ID:

Display Name and Pack ID

Pack ID

The Pack ID:

- Must start with either `cc-` (community-contributed) or `cribl-` (Cribl-contributed). Packs not satisfying this requirement will be immediately rejected.
- Should (by convention) be all lowercase, using the hyphen (`-`) as a delimiter.
- Should **not** include the word "pack."

For Packs normalizing to or from Elastic's Common Schema, the Pack ID (but not the Display Name) should include `-source` or `-dest`.

Display Name

The Display Name is a human-friendly version of the Pack name that:

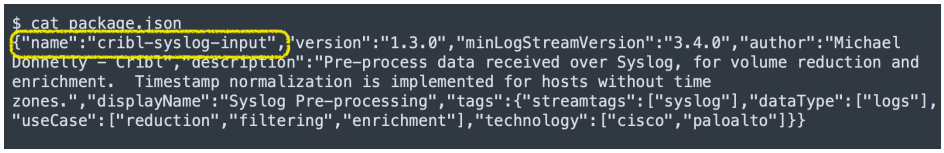
- Omits `cc-` and `cribl-`.
- Uses spaces where appropriate, and proper capitalization.
- If it includes a company or product name, matches the company's preferred spelling, spaces, and capitalization.

You can view or edit the Display Name in the Pack's **Settings** screen.

How to Review the Pack ID

This is the same procedure that the Cribl will perform when reviewing your Pack.

1. Extract the `.crbl` tarball that was submitted to the Dispensary:
   `tar zxvf <your_filename>.crbl`
2. View the `package.json` file:
   `cat package.json`
3. Check the value of the `name` field: this should be the Pack ID that you want.



Verifying the Pack ID

2. Pack Settings

You must review your Pack's settings to verify that the Pack satisfies all of the following criteria.

Display Name

A human-friendly version of the Pack name as defined above.

Version

The version number in the submitted Pack must have a corresponding entry in the README Release Notes.

- The minimum version for initial publication is `0.1.0`. A Pack with a lower version **will be rejected**.
- A release that is considered "beta quality" should be designated as version `0.9.0`.

Description

Use this as a brief "hook" of one to two sentences. Example: `This pack for Syslog inputs will reduce volume, and address timestamp normalization for Syslog senders that omit timezones.`

Author

Community-contributed Packs may include the author's name, a company name, or both.

All three of the following examples are valid:

1. `Jon Smith`
2. `Bazookatron`
3. `Jon Smith - Bazookatron`

Cribl-published Packs must use author's full name, plus `- Cribl`. Example: `Michael Donnelly - Cribl`.

Data Type, Use Cases, Technologies

Select all that apply.

Tags

Tags are optional but recommended, especially where multiple Packs are intended to work together. Tag examples include: `OCSF`, `AWS`.

Logo

The logo is optional, but recommended. You can include a logo associated with the technology addressed by the Pack. For example, a Windows logo for Windows events, an AWS Cloudwatch logo for AWS Cloudwatch data, and so on.

You can set the logo via **Pack Settings** > **Settings** > **Display**.

3. Routes

A Pack submitted must include at least one Route in addition to the default Route.

- A Pack's Routes page must have a Route, with an appropriate filter, routing to a Pipeline.
- The filter should match **only** the set of data that the Pipeline is designed to process. The filter must **not** be set to `true`.
- If the Pack deals with multiple datasets, each dataset should have its own Route/Pipeline pair. (For a good example, look at the PAN Pack.)
- The Pack must not use the default Route to send events to the Pack's custom Pipeline. The default Route's Pipeline may be set to `main`, `devnull`, or `passthru` to handle events that do not match the Pack's filter(s).

- Each Route must include a short description of the dataset being matched.
- Pairs of Routes and Pipelines should have matching names.

4. Pipelines

A Pack submitted must include at least one Pipeline in addition to the stock Pipelines (`devnull`, `main`, `passthru`, `prometheus_metrics`, etc.).

- If the Pack deals with multiple datasets, there must be one Pipeline (and one Route) for each dataset. (For a good example, look at the PAN pack.)
- The Pack must not use the stock Pipelines. Leave these Pipelines unmodified.
- Every pipeline must begin with a Comment Function that gives an overview of the pipeline's operation as a whole. This Comment is a synopsis of what the entire Pipeline will be doing. Refer to the **Syslog Pre-processing** (`cribl-syslog-input`) Pack for an example of the expected level of detail.
- If Function Groups are used within the Pipeline, each Function Group must begin with a Comment Function that gives an overview of what will happen within that Function Group.
- Every Function must have a Description that briefly explains exactly what is happening in that particular Function. During review, change the Pipeline view option to display the Description for each Function.
- Using the sample files provided with the Pack, validate that the Functions work as detailed in the Comment at the top of the Pipeline.

5. Sample Files

The Pack must include at least one sample file for each of its Pipelines.

Sample files:

- Must **not** have an expiration date.
- Must **never** include PII, customer data, or customer hostnames.
- Must be named in such a way that it's obvious which sample file(s) apply to each Pipeline.
- Must contain samples that are adequate for a customer to use to review the operation of a Pipeline. To allow performance analysis of the corresponding Pipeline, a sample should include 20 or more events, with 100 or more recommended.
- May be reused across Pipelines, but it is preferable to include multiple data samples, each specific to a Pipeline's Route and filter.
- Must work as expected on a clean installation of the Pack. This applies to all samples listed in the Pack – Pack developers should test this before submitting the Pack.

6. README

The `README` is the documentation others will use in selecting and deploying the Pack. The `README` requirements ensure that others will be able to deploy the Pack successfully.

The default sections for README, as created when you click **Create Pack**, do not match the required and optional sections in the checklist. You will need to delete or replace certain sections.

In the Pack's **Pack Settings** > **README** screen, validate that the README text is correct. The following rules apply to all sections of the README:

- Proper grammar will help the users of your Pack. Use a grammar and spelling checker to make sure the README is clear and readable.
- If the README includes any text that looks like a placeholder, the Pack will be rejected.
- Trademark or product names must be spelled and capitalized exactly as the company that owns them would dictate. This requirement also applies to Pack settings, inline comments, etc. Examples: Amazon (not amazon), GCP (not Gcp), CrowdStrike (not Crowdstrike).
- The README must include all of the sections listed below (sections **6a** through **6f**). The README may include additional sections as necessary.

6a. About this Pack

Think of this initial section as the first paragraph of an essay, or as the only paragraph a person will read when deciding to use the Pack. It must clearly state what the Pack does. Start with the benefits of using the Pack. In sum, this section should answer the question, **Why should anyone use this Pack?**

If the Pack is tied to a specific data source, dataset, or destination, spell that out clearly here – even if that source or destination is in the Pack name.

6b. Deployment

This **required** section must include all instructions necessary to get the Pack up and running. If the Pack requires any configuration after initial installation - but before it's put it into production - you must specify the configuration steps here.

Filters: If applicable, say what filter must be used to route data to the Pack, or explain how to configure the filter. A filter based on `__inputId` is much more efficient than a match against `_raw`.

Sources: If the Pack requires a specific Source, include information on how to configure that Source. At a minimum, this needs to cover setting up the Source within Cribl. In some cases, this might include

information on how to set up an external environment for collection by Cribl as well.

Destinations: If the Pack requires a specific Destination, include information on how to configure that Destination. Does it need a post-processing Pipeline tied to the Pack?

Pipelines: If the Pack includes Functions (within a Pipeline) that need to tailored, enabled, or disabled, include a full explanation and instructions.

Lookups: If the Pack includes Lookup files that need to be tailored, you must specify the configuration steps here.

6c. Upgrades

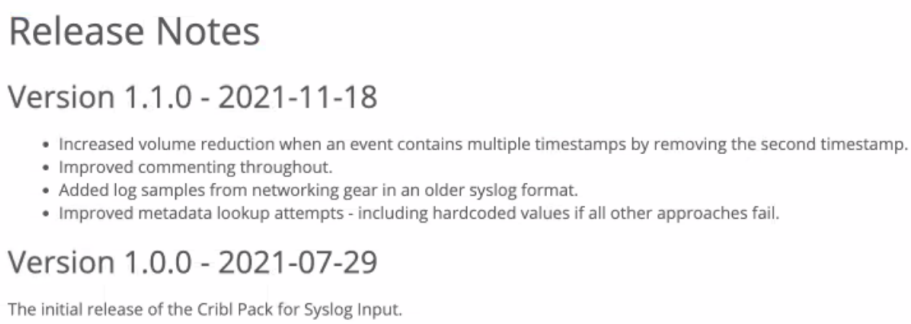
This **optional** section is strongly recommended if you expect others to customize their copy of the Pack and later upgrade the modified Pack.

- If included, this section should cover the usual set of problems with Pack upgrades as described in Upgrading an Existing Pack, and should explain the specifics of upgrading this particular Pack.

6d. Release Notes

This **required** section is a list of release notes where each release note:

- Starts with the version number and date of release, separated by a space, a hyphen, and a space. For example: `Version 1.2.0 - 2022-07-11`.
- Continues with bullet points that describe what has changed and/or is new in the release.



Release notes

Release notes must adhere to the following rules:

- Designate the initial release version `Initial release`.
- The most current version listed must exactly match the Pack version in **Pack Settings** > **Settings**.

- Release notes versions must be listed starting with the most recent (in descending chronological order).

6e. Contributing to the Pack

This **required** section explains how to contribute to the Pack.

- The information provided must be **actionable**. A section reading `[Contributor instructions go here.]` would be rejected.
- Cribl recommends that you refer potential contributors to Cribl's Community Slack. For example, Cribl-created Packs read as follows:

```
To contribute to the Pack, please connect with us on [Cribl Community Slack](h
```

6f. License

This **required** section must include text that designates **and links to** a legitimate license.

- Community-contributed Packs must include a license that allows for use by the general public, such as Apache 2.0, MIT, or GNU.
- Built-by-Cribl Packs must use the following snippet:

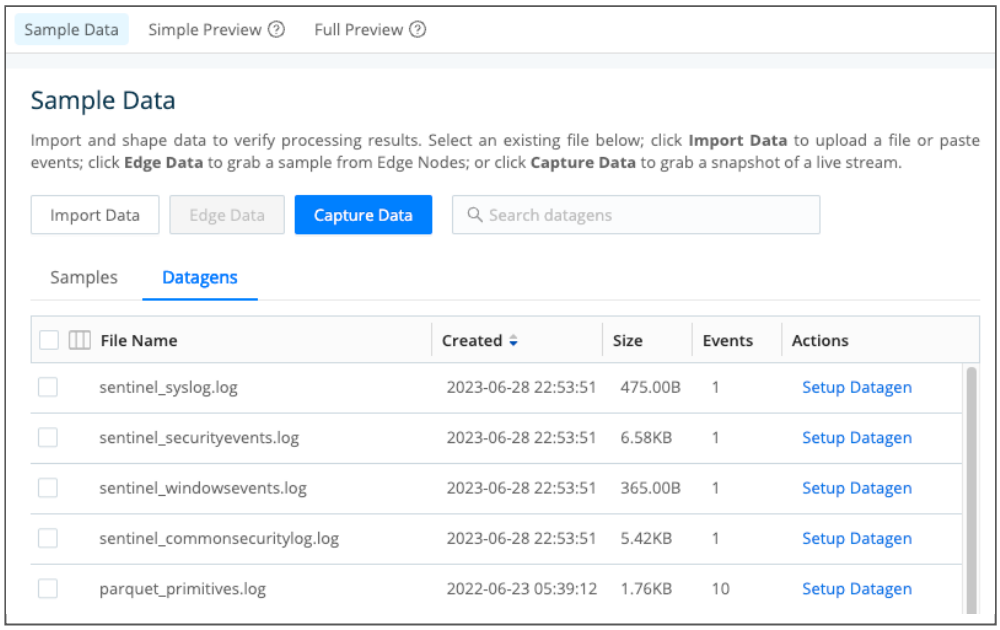
```
This Pack uses the following license: [Apache 2.0](https://github.com/criblio/
```

13.6. Using Datagens

Data generators for testing and troubleshooting

Cribl Edge's datagens feature enables you to generate sample data for the purposes of troubleshooting Routes, Pipelines, Functions, and general connectivity.

Several datagen template files ship with the product, out of the box. You can create others from sample files or live captures.



Preview pane – add samples via paste, attach/upload file, or live capture

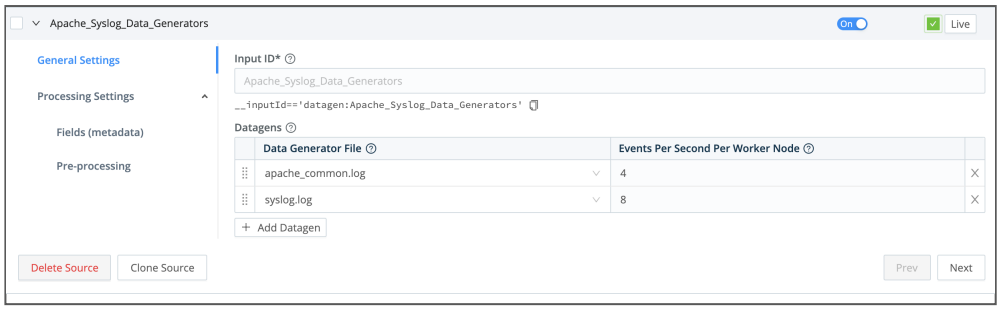
As outlined in the following tutorial: Once you've created a template, you can configure a Datagen Source to use the template to generate real-time data at a given EPS (events per second) rate.

Enabling a Datagen

To see how datagens work, start by enabling a pair of Cribl Edge's out-of-the-box generators:

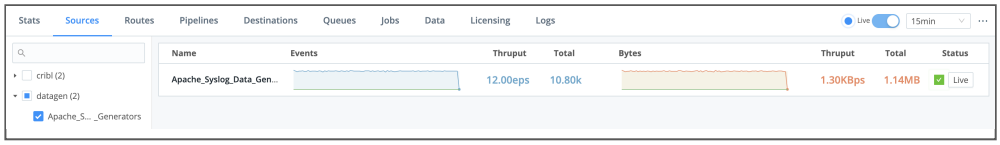
Navigate to **Sources** > **Datagens** and click **New Destination**.

Select a Data Generator File (e.g., `apache_common.log`) and set it at 4 EPS/worker process. Select another Data Generator File (e.g., `syslog.log`) and set it at 8 EPS/worker process. Hit **Save**.



Selecting datagens files and event rates

On the **Monitoring** page, under **Sources**, search for `datagen` and confirm that the Source is generating data.



Creating a Datagen Template from a Sample File

To convert a sample into a template:

Go to **Preview** > **Paste a Sample**, and add a sample like the AWS VPC Flow logs below:

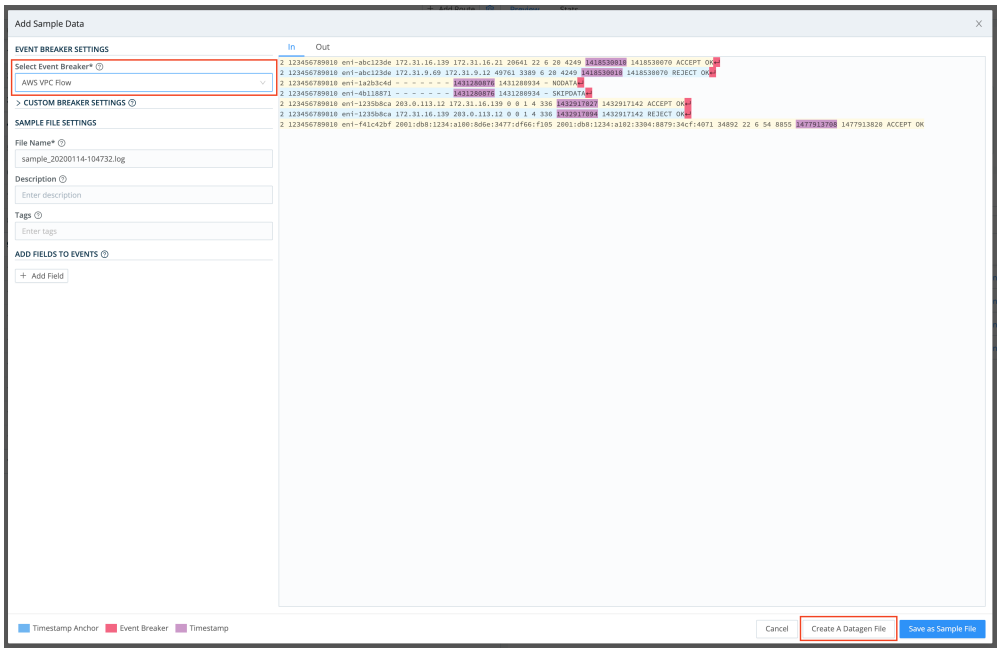
Sample VPC Flow Logs

```
2 123456789010 eni-abc123de 172.31.16.139 172.31.16.21 20641 22 6 20 4249 14185300
2 123456789010 eni-abc123de 172.31.9.69 172.31.9.12 49761 3389 6 20 4249 1418530010
2 123456789010 eni-1a2b3c4d - - - - - - - 1431280876 1431280934 - NODATA
2 123456789010 eni-4b118871 - - - - - - - 1431280876 1431280934 - SKIPDATA
2 123456789010 eni-1235b8ca 203.0.113.12 172.31.16.139 0 0 1 4 336 1432917027 1432
2 123456789010 eni-1235b8ca 172.31.16.139 203.0.113.12 0 0 1 4 336 1432917094 1432
2 123456789010 eni-f41c42bf 2001:db8:1234:a100:8d6e:3477:df66:f105 2001:db8:1234:a
```

From the **Event Breaker** drop-down, select **AWS VPC Flow** to ensure that:

- The pasted text gets broken properly into individual events (notice the Event Breaker on newlines).
- Timestamps are extracted correctly (text highlighted purple below).

Once you've verified these results, click **Create a Datagen File**.

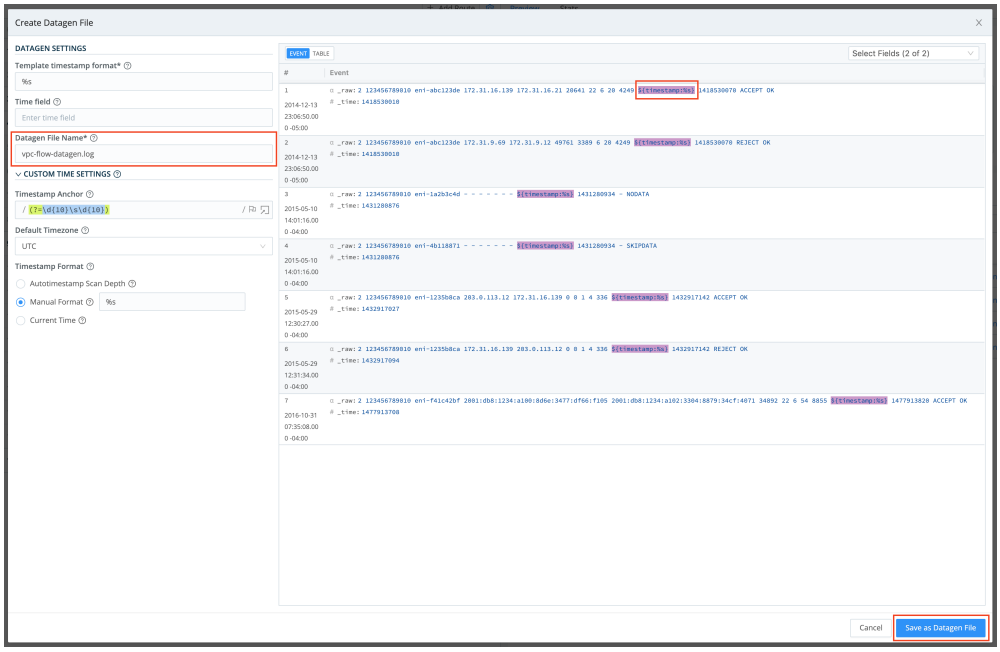


Creating a datagen template

On the resulting **Create Datagen File** screen:

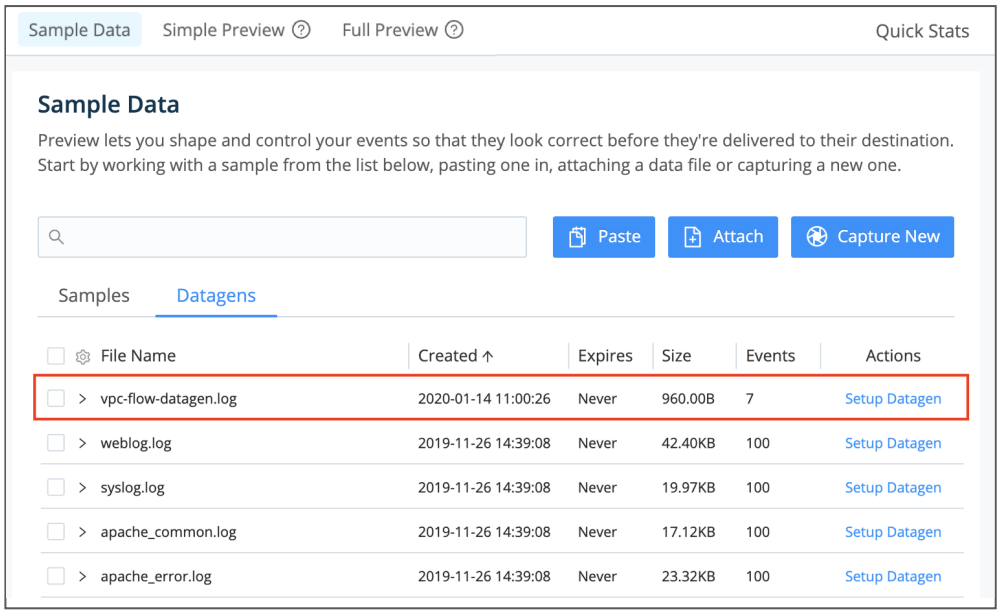
- Enter a file name, e.g.: `vpc-flow-datagen.log`
- Ensure that the timestamp template format is correct: `${timestamp: %s}`
  `${timestamp: <format>}` is a template that the datagen engine uses to insert the current time – in each newly generated event – using the given format. In this case, `%s` is the desired `strftime` format for the timestamp (i.e., the epoch).

Once you've verified these results, click **Save as Datagen File**.



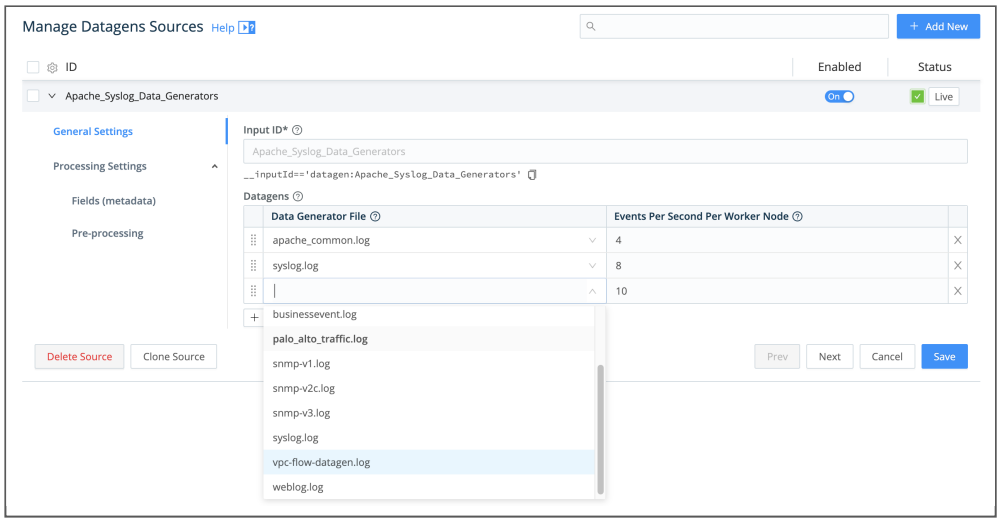
Saving a named datagen template

To confirm that the datagen file has been created, check **Preview** > **Datagens**.



Verifying datagen file creation

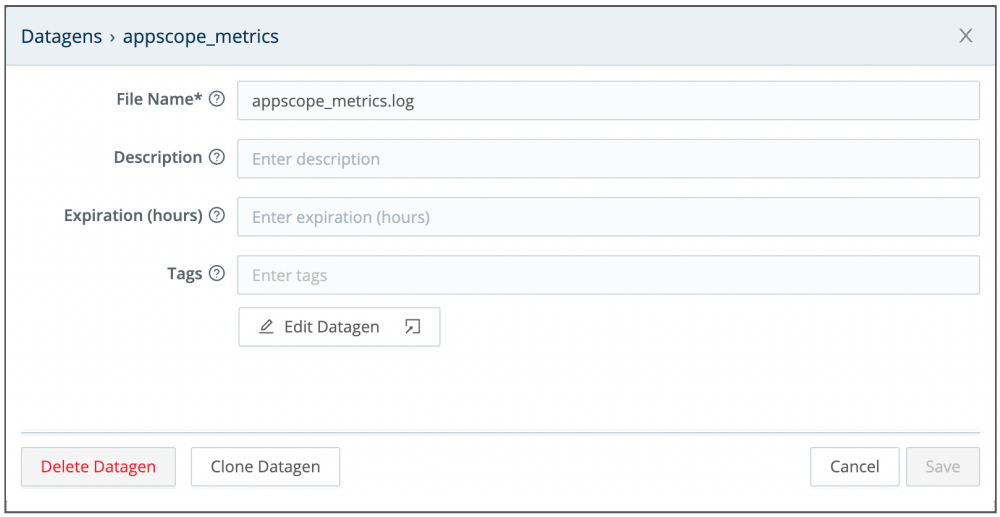
Now, to start using your newly created datagen file, go back to **Sources** > **Datagens**. Add it using the drop-down shown below.



Adding new template file to datagens Source

Modifying a Datagen

1. In the right Preview pane, select the **Datagens** tab.
2. Hover over the file name you want to modify. This displays an edit (pencil) button to its left.
3. Click that button to open the modal shown below. It provides options to edit the datagen, clone it, delete it, or modify its metadata (**File name**, **Description**, **Expiration time**, and **Tags**).



Options for modifying a datagen

4. To make changes to the datagen, click the modal's **Edit Datagen** button. This opens the **Edit Datagen** modal shown below, exposing the raw data that this datagen uses to generate events.
5. Edit the raw data as desired.
6. Click **Update Datagen** to resave the modified datagen, or click **Save as New Datagen** to give the modified version a new name.

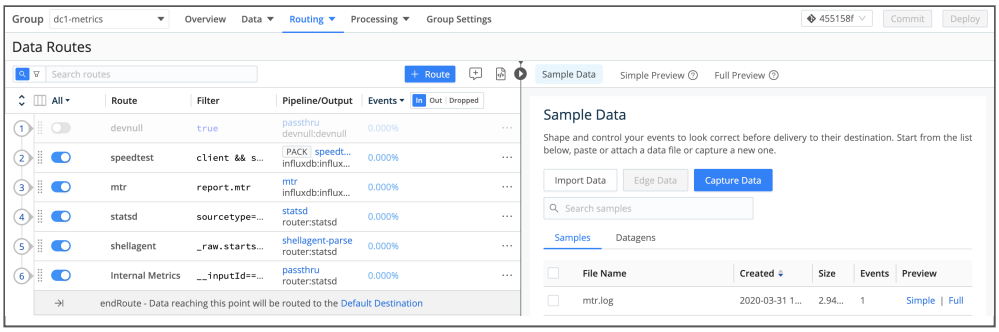
Editing a datagen

13.7. Data Preview

Cribl Edge's Sample Data Preview features enable you to visually inspect events as they flow into and out of a Pipeline. Preview helps you shape and control events before they're delivered to a Destination, and helps you troubleshoot Pipeline Functions.

Preview works by taking a set of sample events and passing them through the Pipeline, while displaying the inbound and outbound results in a separate pane. Any time a Function is modified, added, or removed, the Pipeline changes, and so does its displayed output.

The Preview pane is shown below, to the right of the Pipelines pane.

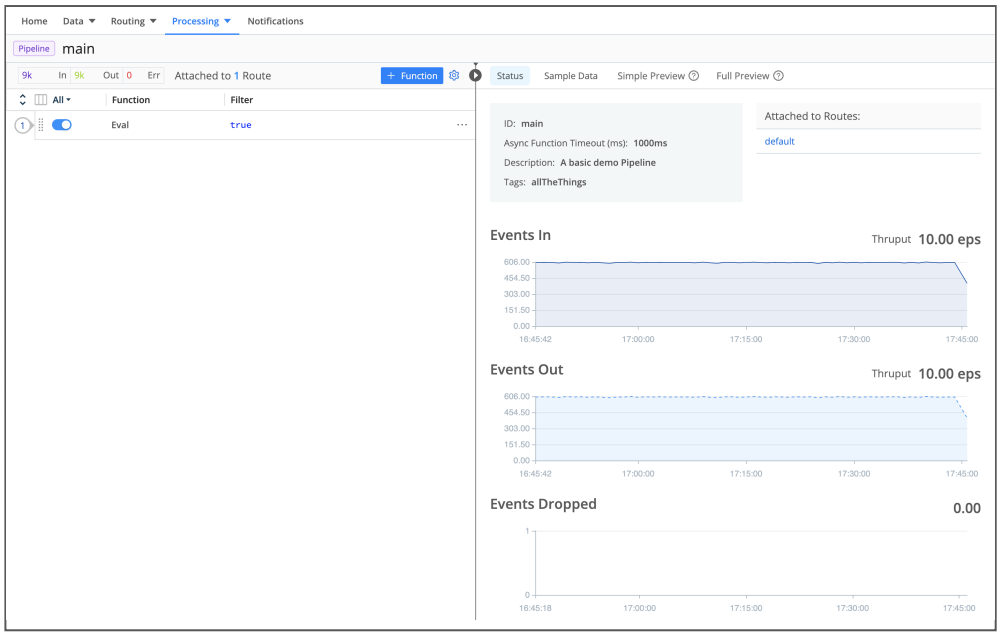


Preview options

You can press the] (right-bracket) shortcut key to toggle the visibility of the Preview pane. (This shortcut works when no field has focus.)

Checking Pipeline Status

From the Pipelines page, click any Pipeline at left to display a **Status** tab in the right Preview pane. Click this tab to reveal graphs of the Pipeline's **Events In**, **Events Out**, and **Events Dropped** over time, along with summary throughput statistics (in events per second).



Pipeline status tab

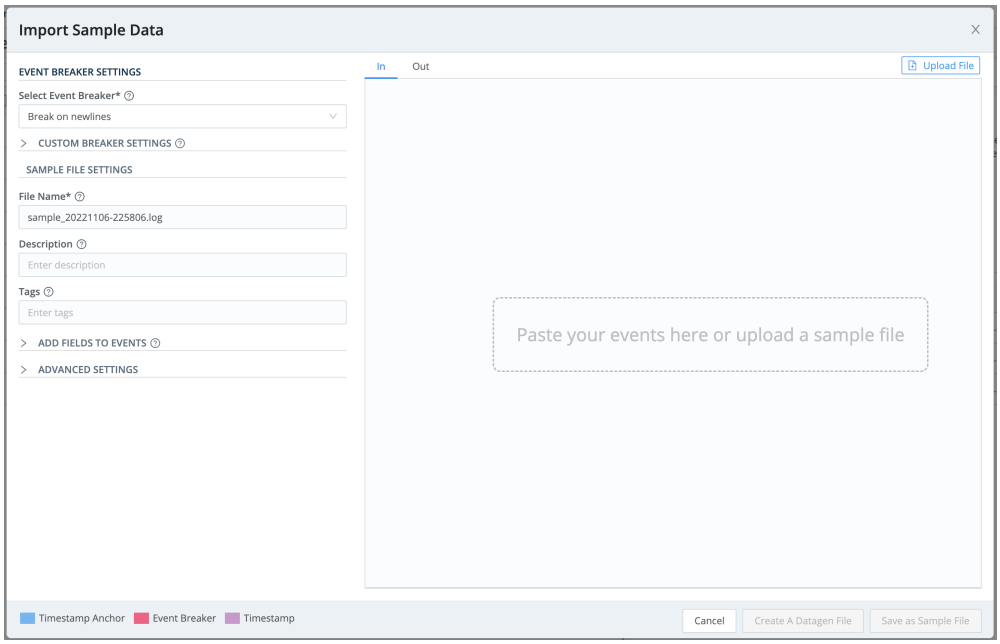
Adding Sample Data

When you're on the Pipelines or Routes page, you can add samples through any of the Preview pane's supported options: **Import Data** (which opens a modal where you can paste data from your clipboard or use the **Upload File** button); **Edge Data** (where you can import data from any Edge Nodes deployed on your account); and **Capture New**. The **Import Data** options work with content that needs to be broken into events, while the **Capture New** option works with events only.

The **Edge Nodes** option requires at least one working Edge Node, and is not available when you've teleported to the Node.

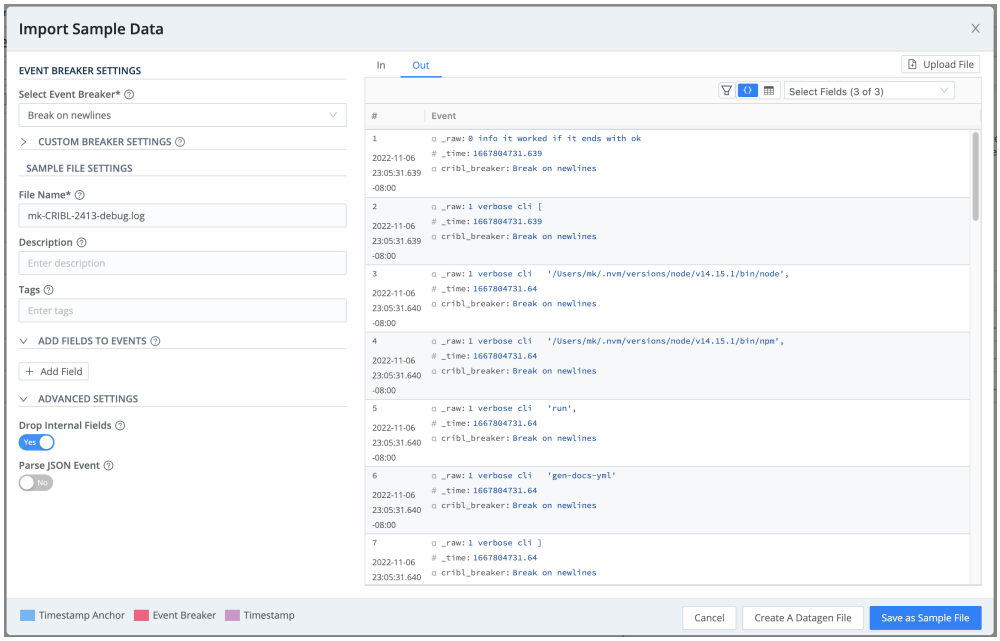
Paste Area

Once you've clicked the **Import Data** button, or imported Edge data, you'll see an **Import Sample Data** modal, where you can paste data or upload a file.



Import Sample Data modal

Once your data is onboard, you have options to modify it using Event Breakers, to add and drop fields, and to save it to a file.

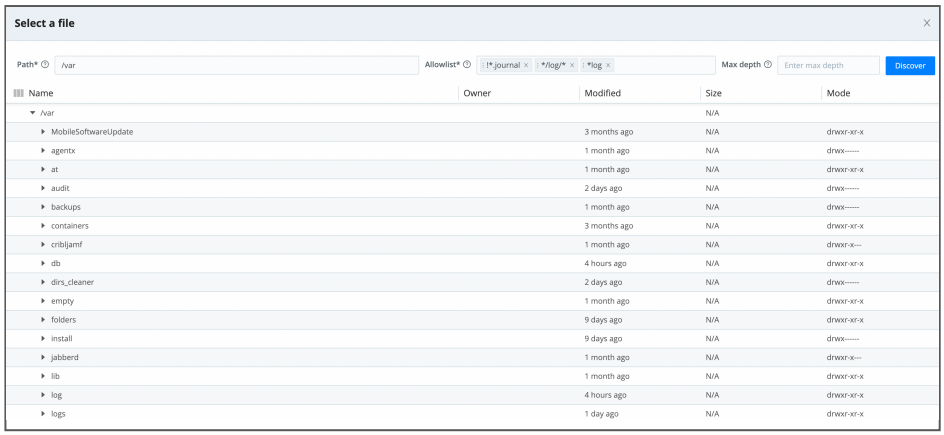


Working with imported data

Edge Data

To upload data from a file on an Edge node:

1. Click the **Edge Data** button, and navigate to the Edge node where the file is stored. Clicking the Edge Node opens the **Select a file** modal, shown below.



Select a file modal

2. Use the available filters to narrow the results:
    - **Path**: Sets the location from which to discover files.
    - **Allowlist**: This filter supports wildcard syntax (as shown in the screenshot above), and supports the exclamation mark (!) for negation.
    - **Max depth**: Sets which layers of files to return highlighted in bold typeface. By default, this field is empty, which implicitly specifies `0`. This default will boldface only the top-level files within the **Path**.
3. Once you find the file you want, click its name to add its contents to the **Add Sample Data** modal, where you'll finish configuring the data sample.

Event Breaker Settings

Cribl Event Breakers are regular expressions that tell Cribl Edge how to break the file or pasted content into events. Breaking will occur at the **start** of the match. Cribl Edge ships with several common breaker patterns out of the box, but you can also configure custom breakers. The UI here is interactive, and you can iterate until you find the exact pattern.

Troubleshooting Event Breakers

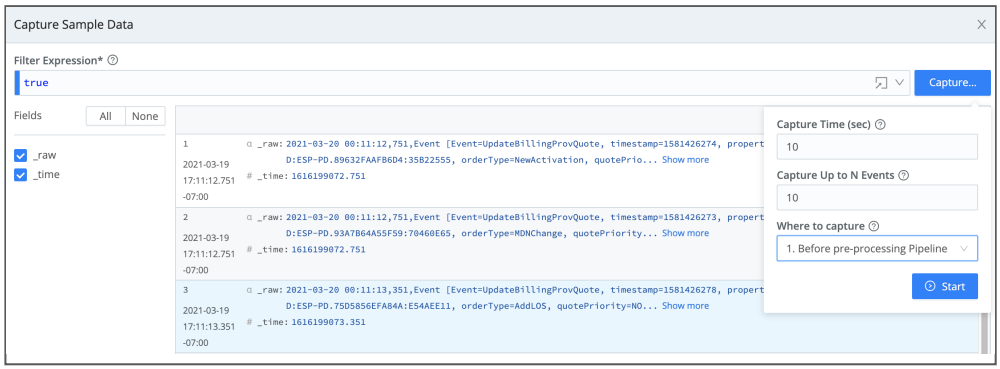
If you notice fragmented events, check whether Cribl Edge has added a `__timeoutFlush` internal field to them. This diagnostic field's presence indicates that the events were flushed because the Event Breaker buffer timed out while processing them. These timeouts can be due to large incoming events, backpressure, or other causes.

Capturing Sample Data

The **Capture Data** button opens a slightly different modal – it does not require event breaking.

In order to capture live data, you must have Edge Nodes registered to the Fleet for which you're viewing events. You can view registered Edge Nodes from the **Status** tab in the Source.

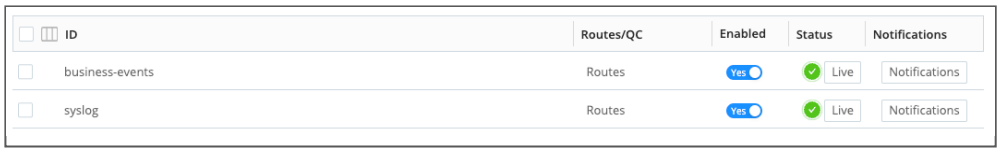
In the composite screenshot below, we've already captured some events using the **Capture** drop-down.



Capture Data > Capture Sample Data modal

Capturing from a Single Source or Destination

To capture data from a single enabled Source or Destination, it's fastest to use the Sources or Destinations UI instead of the Preview pane. You can initiate an immediate capture by clicking the **Live** button on the Source's or Destination's configuration row.

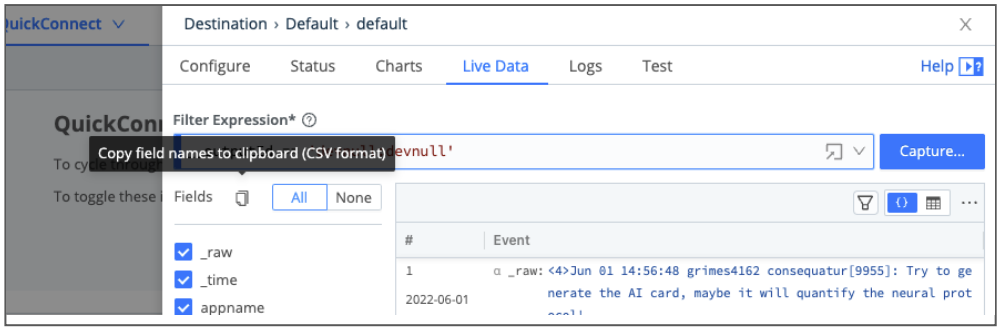


Source > Live button

You can similarly start an immediate capture from within an enabled Source's or Destination's configuration modal, by clicking the modal's **Live Data** tab.

Destination modal > Live Data tab

Beside the **Live Data** tab's **Fields** selectors is a Copy button, which enables you to copy the field names to the clipboard in CSV format. The **Logs** tab also provides this copy button.



Destination modal > Live Data - Copy Fields Icon

Controlling Sample Size

To prevent in-memory samples from getting unreasonably large, samples that you input by any means (Import Data, Edge Data, or Capture Data) are constrained by a limit set as follows:

- On a single instance at **Settings** > **General Settings > Limits** > **Storage** > **Max sample size**.
- On distributed Worker Groups at **Manage** > `<group-name>` > **Fleet Settings** > **General Settings** > **Limits** > **Storage Max sample size**.

In each location, the default limit is `256 KB`. You can adjust this upward (to a maximum of `3 MB`) or downward.

Very Large Integer Values

Cribl Edge's JavaScript implementation can safely represent integers only up to the Number.MAX_SAFE_INTEGER constant of about 9 quadrillion (precisely, {2^53}-1). Data Preview will round down any integer larger than this, and trailing 0's might indicate such rounding.

Fields

In the **Capture Data** and **Live Data** modals, use the **Fields** sidebar (at left) to streamline how events are displayed. You can toggle among **All** fields, **None** (to reset the display), and check boxes that enable/disable individual fields by name.

Field Type Symbols

Within the right Preview pane, each field's type is indicated by one of these leading symbols:

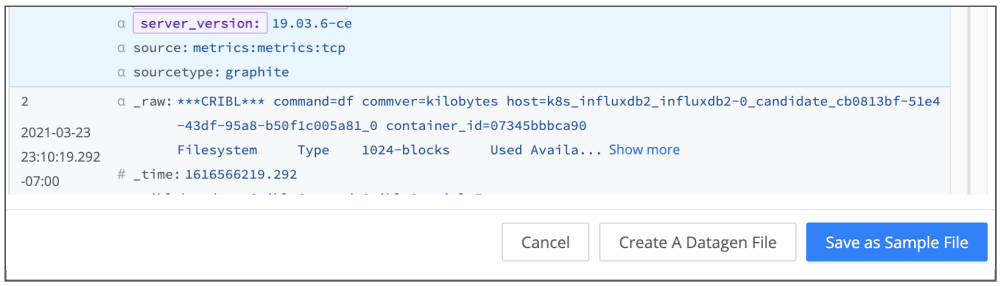
Symbol	Meaning
α	string
#	numeric
b	boolean
m	metric
{}	JSON object
[]	array

On JSON objects and arrays, you'll also see:

Symbol	Meaning
+	expandable
-	collapsible

Saving Sample Data

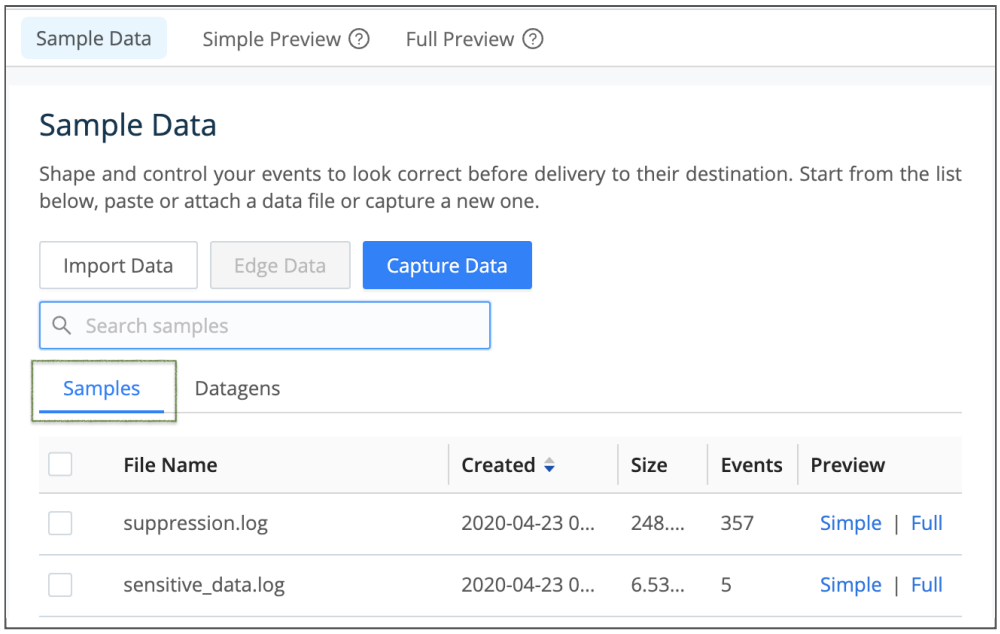
The Preview pane's **Import Data** or **Capture Data** modal, once you've successfully populated it with data, provides options to save the data as a sample and/or datagen file. Click the appropriate button, accept or modify the default/generated file name and other options, and confirm the save.



Saving sample data

Accessing and Managing Data Files

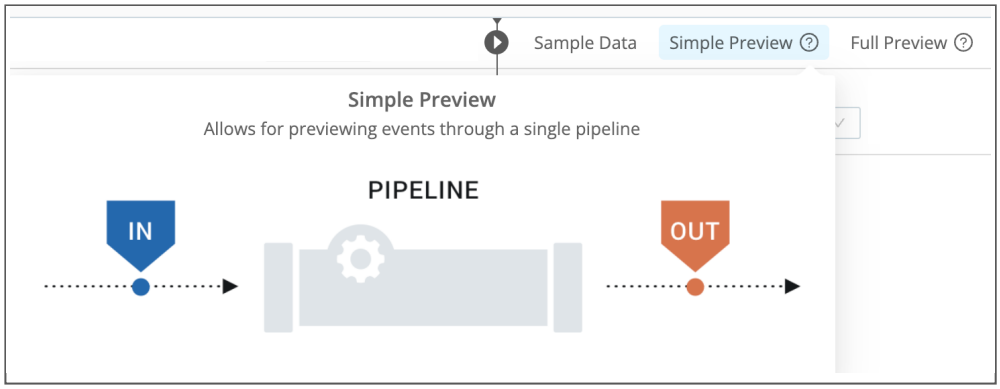
As you add more samples to your system, you can easily access them via the **Sample data file** drop-down. You can also manage and modify sample files via the **Samples** tab highlighted below.



Managing sample files

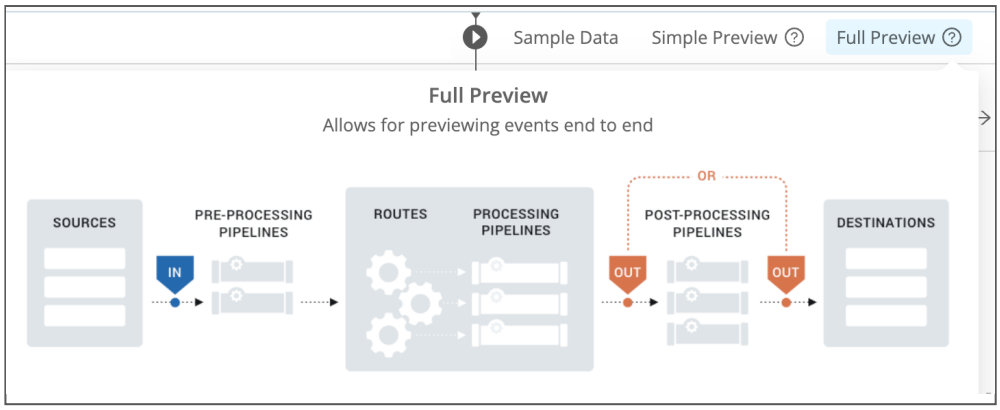
Simple Versus Full Preview

Click **Simple** or **Full** beside a file name to display its events in the Preview pane. The **Simple Preview** option enables you to view events on either the **IN** or the **OUT** (processed) side of a single Pipeline.



Simple Preview schematic

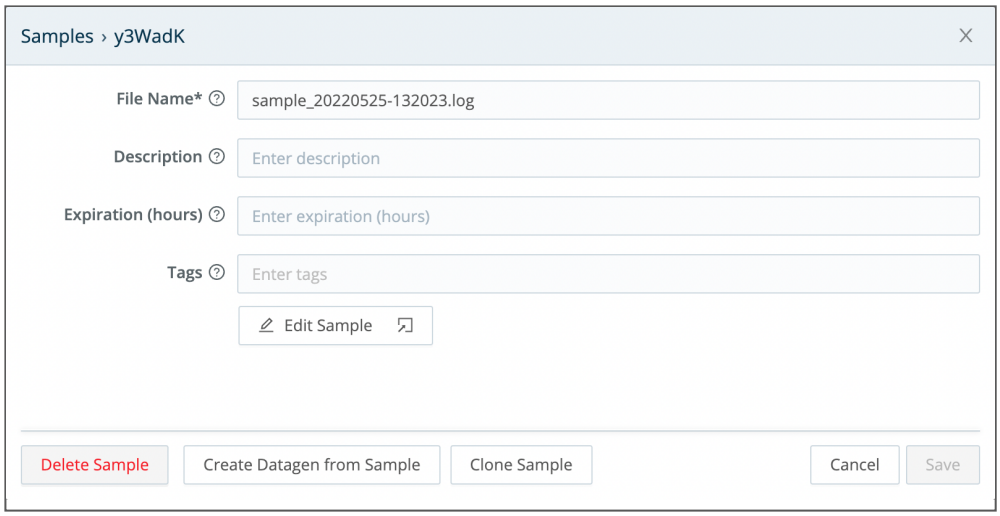
The **Full Preview** option gives you a choice of viewing events on the **OUT** side of either the processing or post-processing Pipeline. Selecting this option expands the Preview pane's upper controls to include an **Exit Point** drop-down, where you make this choice.



Full Preview schematic

Modifying Sample Files

1. In the right Preview pane, select the **Samples** tab.
2. Hover over the file name you want to modify. This displays an edit (pencil) button to its left.
3. Click that button to open the modal shown below. It provides options to edit, clone, or delete the sample, save it as a datagen Source, or modify its metadata (**File name**, **Description**, **Expiration time**, and **Tags**).



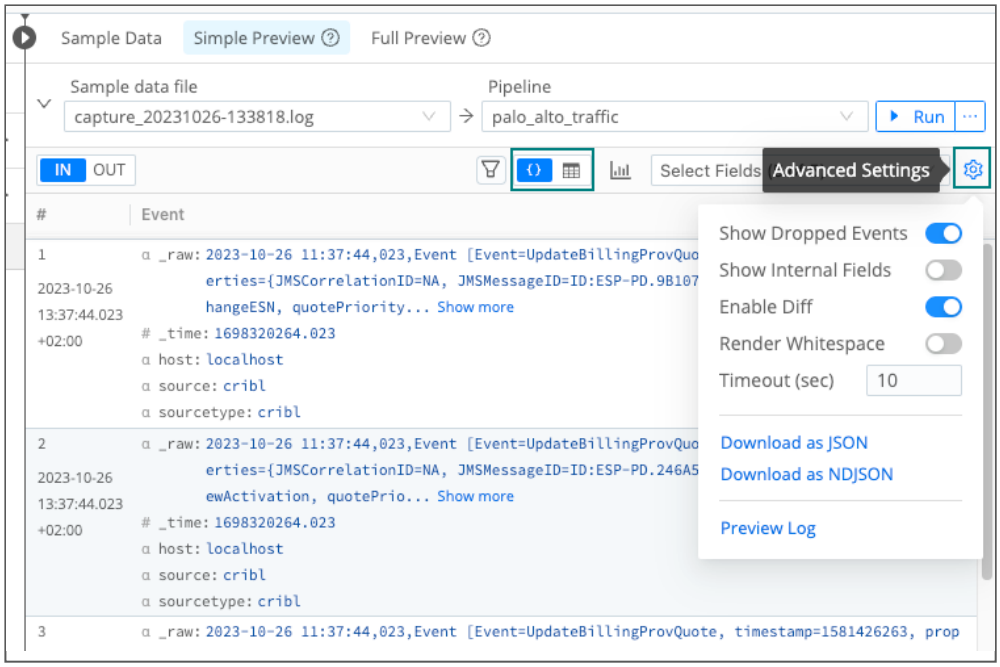
Options for modifying a sample

4. To make changes to the sample, click the modal's **Edit Sample** button. This opens the **Edit Sample** modal shown below, exposing the sample's raw data.
5. Edit the raw data as desired.
6. Click **Update Sample** to resave the modified sample, or click **Save as New Sample** to give the modified version a new name.

Editing a sample file

IN Tab: Displaying Samples on the Way IN to the Pipeline

The Preview pane offers two display options for events: Event and Table. Each format can be useful, depending on the type of data you are previewing. This screenshot shows Event view:



Event, Table, and Advanced options (composite screenshot)

On the **Advanced Settings** menu at the upper right, the first few toggles are self-explanatory, and are used primarily to filter the OUT tab's display of processed data. The following subsections cover the less-obvious controls at the menu's bottom.

Render Whitespace

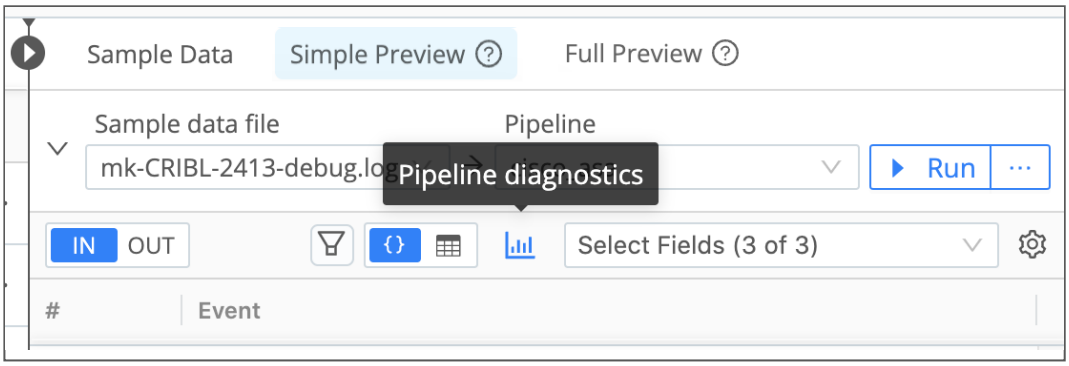
This toggles between displaying carriage returns, newlines, tabs, and spaces as white space, versus as(respectively) the symbols ␍, ␊, →, and ·.

Timeout (Sec)

If large sample files time out before they fully load, increase this field's default value of `10` seconds. A blank field is interpreted as the minimum allowed timeout value of `1` second.

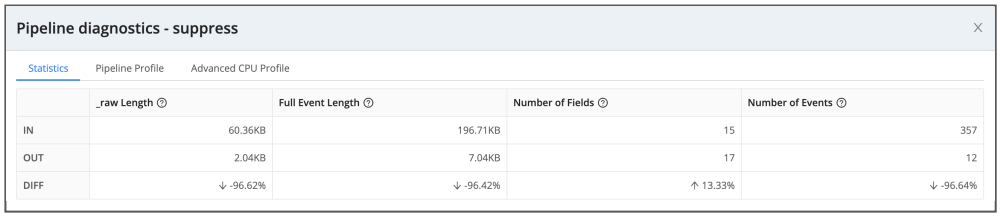
Pipeline Profiling

At the Preview pane's top, you can select a sample data file and Pipeline, then click the Pipeline diagnostics (bar-graph) button to access statistics on the selected Pipeline's performance. (The same button is available on Pipelines within a Pack.)



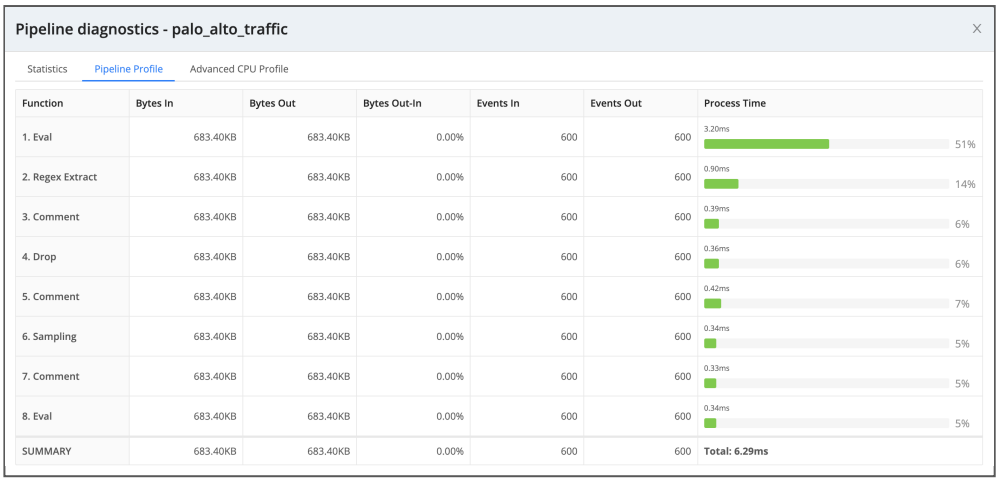
Pipeline diagnostics

The resulting modal's **Statistics** tab displays basic stats about the Pipeline's processing impact on field and event length and counts.



Pipeline diagnostics > Statistics tab

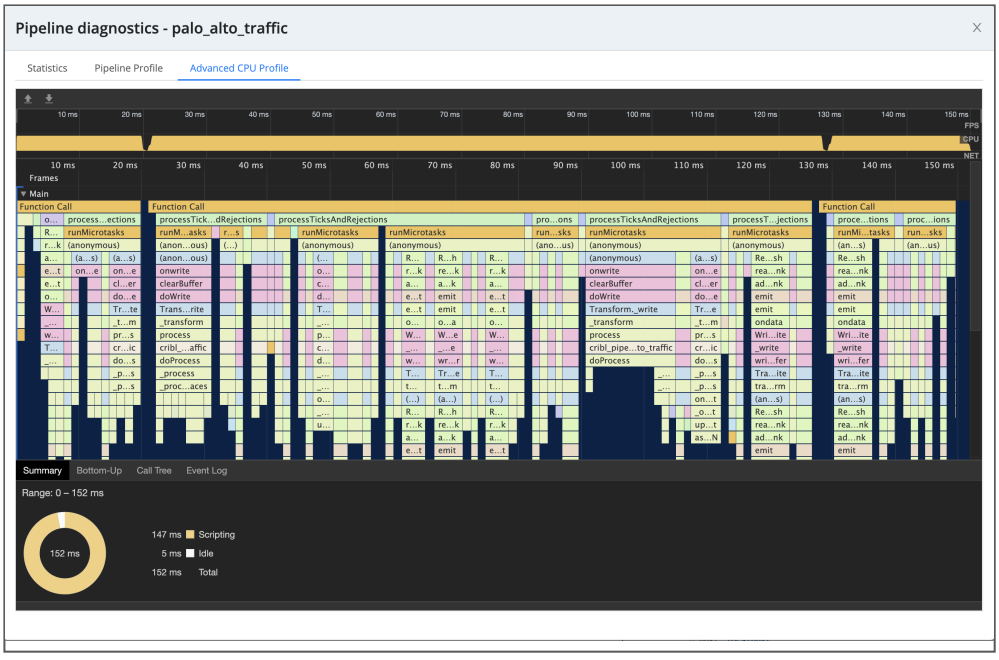
The **Pipeline Profile** tab, available from the **Simple Preview** tab, helps you debug your Pipeline by showing individual Functions' contributions to data volume, events count, and the Pipeline's processing time. Every Function has a minimum processing time, including Functions that make no changes to event data, such as the Comment Function.



Pipeline diagnostics > Pipeline Profile tab

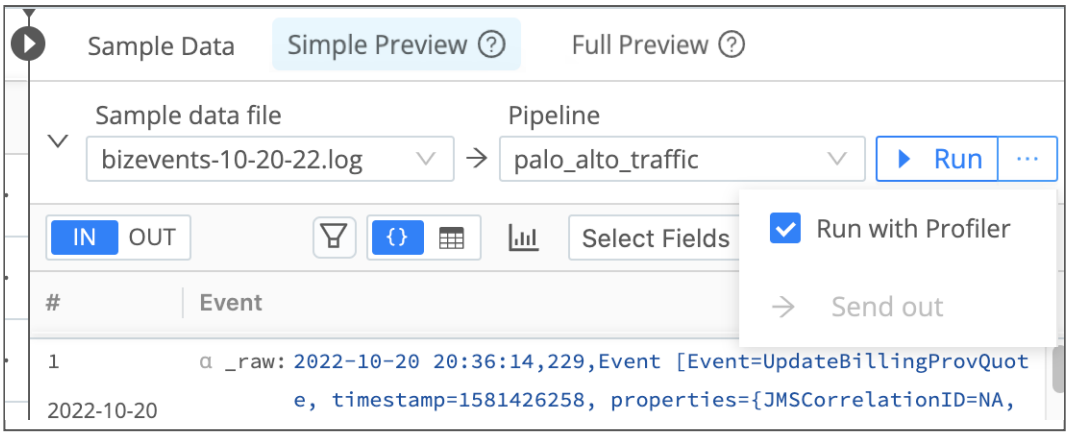
When you profile a Chain Function, this tab condenses all of the chained Pipeline's processing into a single **Chain** row. To see individual statistics on the aggregated Functions, profile the chained Pipeline separately.

The **Advanced CPU Profile** tab, available from the **Simple Preview** tab, displays richer details on individual Function stacks.



Pipeline diagnostics > Advanced CPU Profile tab

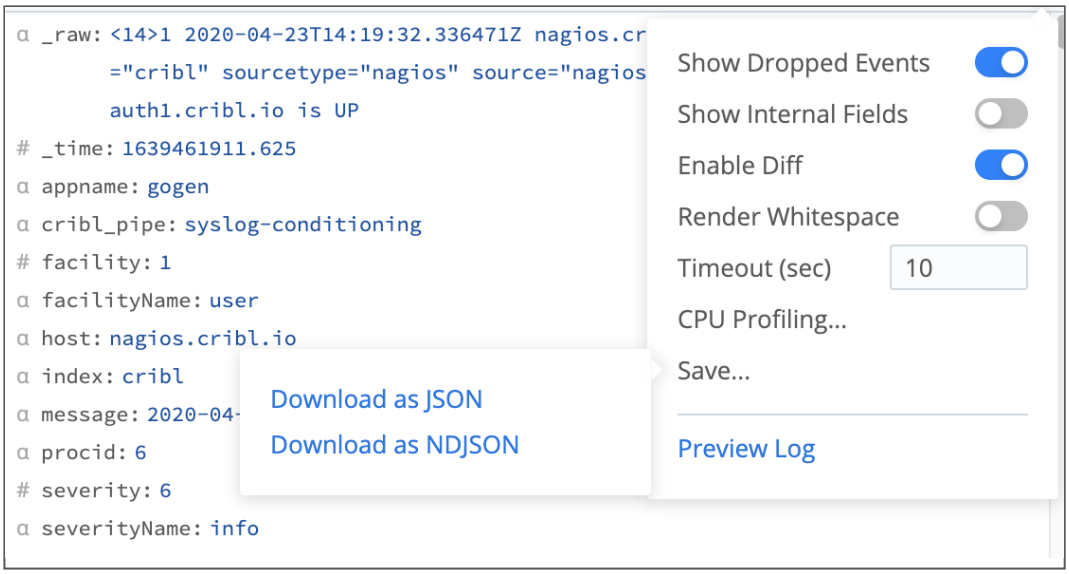
You can disable the profiling features – and regain their own (minimal) resources – by clicking the drop-down list beside the **Run** button and clearing the **Run with Profiler** check box.



Disabling Pipeline profiling

Save

The **Save** submenu enables you to save your captured data to a file, using either the **Download as JSON** or the **Downoad as NDJSON** (Newline-Delimied JSON) option.



Saving sample data as JSON

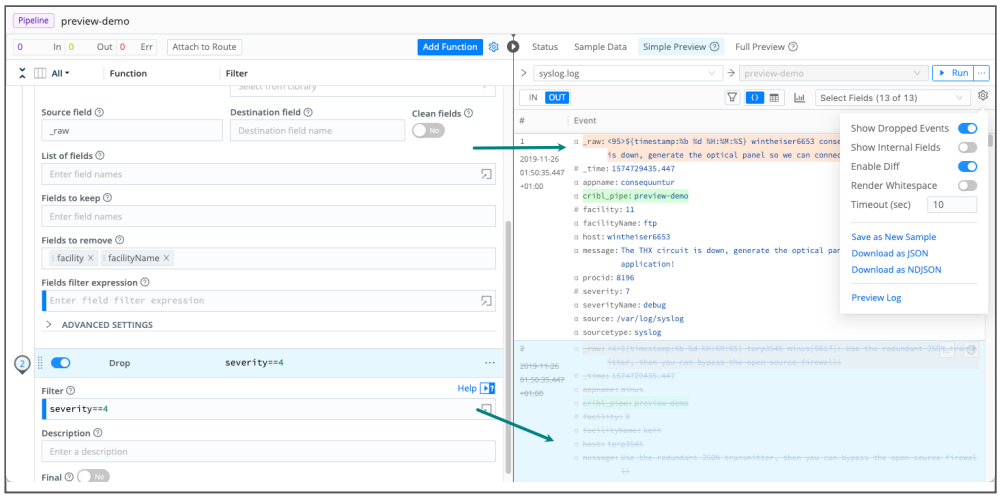
Preview Log

The final option on the **Advanced Settings** menu opens a modal where you can preview Cribl Edge's internal logs summarizing how this data sample was processed and captured.

OUT Tab: Displaying Samples on the Way OUT of the Pipeline

As data traverses through a Pipeline's Functions, events can be modified, and some might be dropped altogether. The **OUT** tab indicates changes using this color coding:

- **Dropped events**: When events are dropped, the **OUT** tab displays them as grayed-out text, with strikethrough. You can control their display using the **Advanced Settings** menu's **Show Dropped Events** toggle.
- **Added fields**: When Cribl Edge's processing adds new fields, these fields are highlighted green. You can control these fields' display using the **Select Fields** drop-down.
- **Redacted fields**: These fields are highlighted amber.
- **Deleted fields**: These fields are highlighted red.



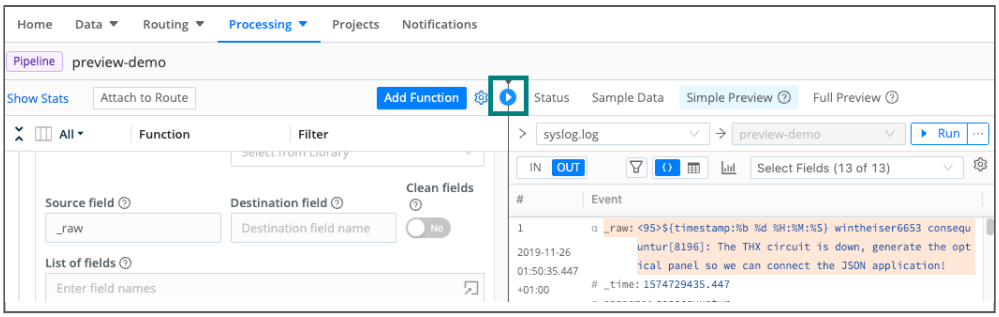
Dropped and added fields in a Pipeline's output

The **OUT** tab displays the same Event versus Table buttons as the IN tab. It also displays the same **Advanced Settings** menu options – and here, you can use the menu's **Show Dropped Events**, **Show Internal Fields**, and **Enable Diff** toggles to clarify how the data has been transformed by the Pipeline.

Enable **Show Internal Fields** to discover fields that Cribl Edge adds to events, as well as Source-specific fields that Cribl Edge forwards from upstream senders.

Managing the Preview Pane

On the pane divider between Route or Pipeline and the preview, click Collapse to hide the Preview pane. This allows the Route or Pipeline configuration to expand to your browser's full width. (The Preview pane collapses automatically on narrow viewports.)



Collapse / Expand toggle

Click Expand at your browser's right edge to restore the split view. The pane divider will snap back to wherever you last dragged it.

13.8. Data Onboarding

Onboarding data into Cribl Edge can vary in complexity, depending on your organization's needs, requirements, and constraints. Proper onboarding from all Sources is key to system performance, troubleshooting, and ultimately the quality of data and decisions both in Cribl Edge and in downstream Destinations.

General Onboarding Steps

Typically, a data onboarding process revolves around these steps, both before and after turning on the Source:

- Create configuration settings.
- Verify that settings do the right thing.
- Iterate.

Below, we break down individual steps.

Before Turning On the Source

Cribl recommends that you take the following steps to verify and tune incoming data, before it starts flowing.

Preview Sample Data

Use a sample of your real data in Data Preview. Sample data can come from a sample Source file that you upload or paste into Cribl Edge.

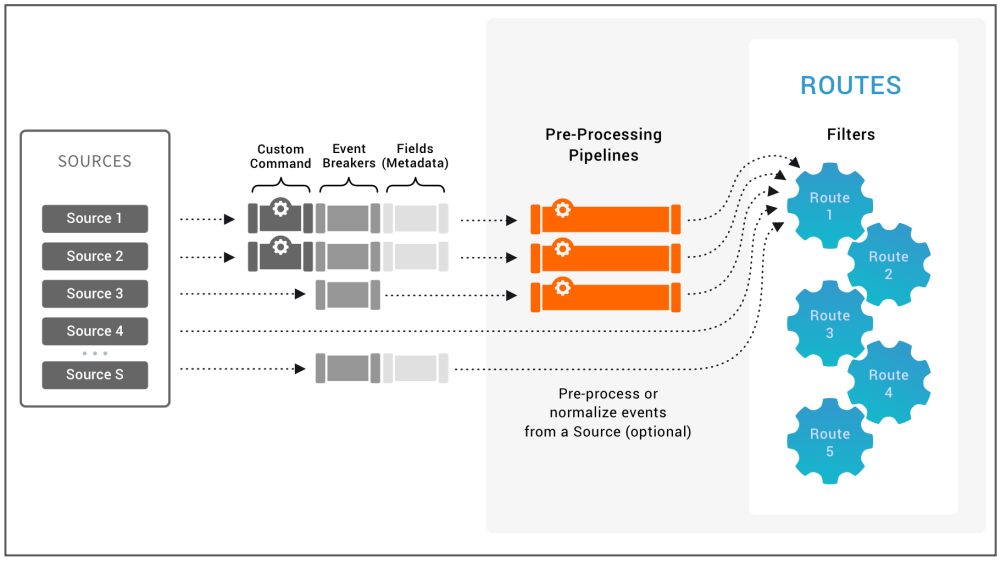
You can also obtain sample data in a live data capture from a Source. One way to do this **before** going to production is to configure your Source with a **devnull** Pipeline (which just drops all events) as a pre-processing Pipeline. Then, let data flow in for just long enough to capture a sufficient sample.

Very Large Integer Values

Cribl Edge's JavaScript implementation can safely represent integers only up to the Number.MAX_SAFE_INTEGER constant of about 9 quadrillion (precisely, {2^53}-1). Data Preview will round down any integer larger than this, and trailing 0's might indicate such rounding down.

Check the Processing Order

While events can be processed almost arbitrarily by Functions in Cribl Edge Pipelines, make sure you understand the event processing order. This is very important, as it tells you exactly where certain processing steps occur. For instance, as we'll see just below, quite a few steps can be accomplished at the Source level, before data even hits Cribl Edge Routes.



Source-level processing options

Custom Command

Where supported, data streams will be handled by **custom commands**. These are external system commands that can (optionally) be used to pre-process the data. You can specify any command, script, etc., that consumes via `stdin` and outputs via `stdout`.

Verify that such commands are doing what's expected, as they are the very **first** in a series of processing steps.

Event Breakers

Next, data streams are handled by Event Breakers, which:

- Convert data streams into discrete events.
- Extract and assign timestamps to each event.

If the resulting events do not look correct, feel free to use **non-default** breaking rules and timestamp recognition patterns. Downstream, you can use the Auto Timestamp Function to modify `_time` as needed, if timestamps were not recognized properly. Examples of such errors are:

- Timestamps too far out in the future or past
- Wrong timezone.
- Incorrect timestamp is selected from multiple timestamps present in the event.

Fields

Next, events can be enriched with Fields . This is where you'd add static or dynamic fields to all events delivered by a particular Source.

Pre-Processing Pipeline

Next, you can optionally configure a pre-processing Pipeline on a particular Source. This is extremely useful in these cases:

- Drop non-useful events as early as possible (so as to save on CPU processing).
- Normalize events from this Source to conform a certain shape or structure.
- Fix/touch up events accordingly. E.g., if event breakers assigned the wrong timestamp, this is the best place to use the Auto Timestamp Function to adjust `_time`.

We Can't Say This Enough

Verify, verify, verify, data integrity before turning on the Source.

After Turning On the Source

Use data Destinations to verify that certain metrics of interest are accurate. This will depend significantly on the capabilities of each Destination, but here's a basic checklist of things to ensure:

- Timestamps are correct.
- All necessary fields are assigned to events.
- All expected events show up correctly. (E.g., if a Drop or Suppress Function was configured, ensure that it's not dropping unintended events.)
- Throughput – both in bytes and in events per second (EPS) – is what's expected, or is within a certain tolerance.

Iterate

Iterate on the steps above as necessary. E.g., adjust fields values and timestamps as needed.

Remember that there is almost always a workaround. Any arbitrary event transformation that you need is likely just a Function or two away.

14. Functions

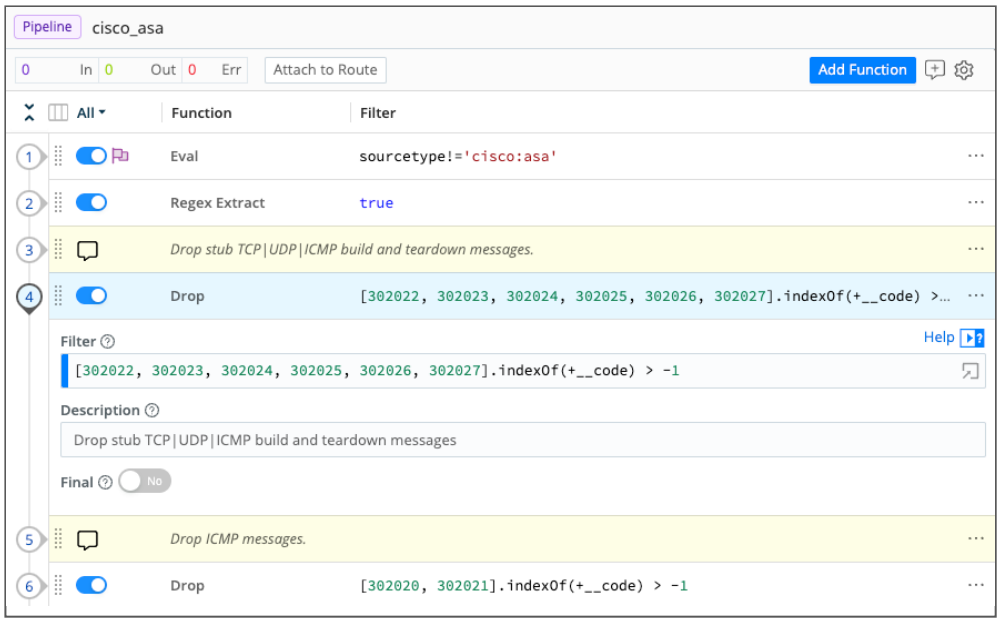
When events enter a Pipeline, they're processed by a series of Functions. At its core, a Function is code that executes on an event, and it encapsulates the smallest amount of processing that can happen to that event.

The term "processing" means a variety of possible options: string replacement, obfuscation, encryption, event-to-metrics conversions, etc. For example, a Pipeline can be composed of several Functions – one that replaces the term `foo` with `bar`, another one that hashes `bar`, and a final one that adds a field (say, `dc=jfk-42`) to any event that matches `source=='us-nyc-application.log'`.

How Do They Work

Functions are atomic pieces of JavaScript code that are invoked on each event that passes through them. To help improve performance, Functions can be configured with filters to further scope their invocation to matching events only.

You can add as many Functions in a Pipeline as necessary, though the more you have, the longer it will take each event to pass through. Also, you can enable and disable Functions within a Pipeline as necessary. This lets you preserve structure as you optimize or debug.



Functions stack in a Pipeline

You can reposition Functions up or down the Pipeline stack to adjust their execution order. Use a Function's left grab handle to drag and drop it into place.

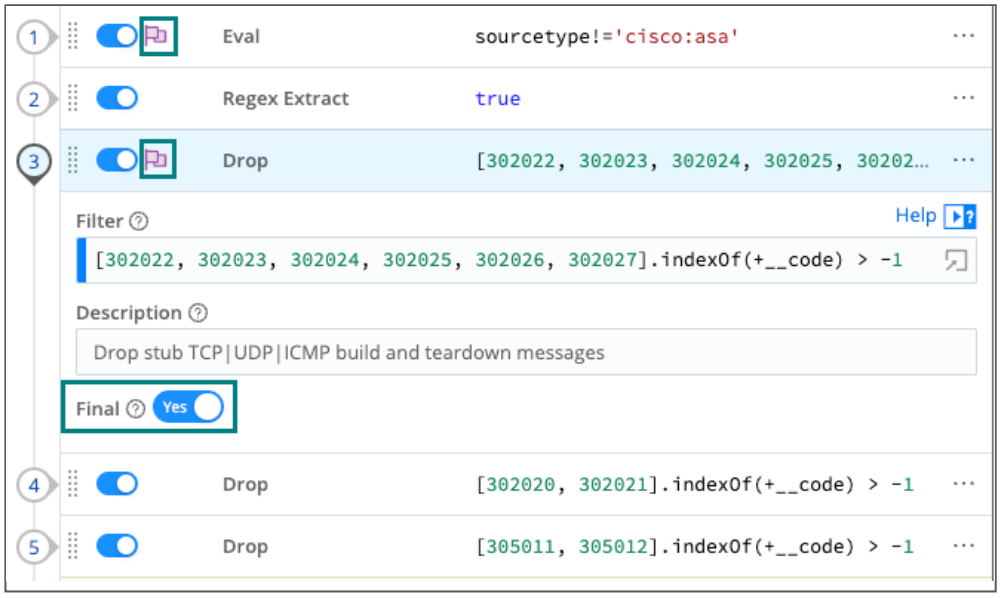
The `Final` Toggle

The `Final` toggle in Function settings controls what happens to the results of a Function.

When `Final` is toggled to `No` (default), events will be processed by this Function, **and** then passed on to the next Function below.

When `Final` is toggled to `Yes`, the Function "consumes" the results, meaning they will not pass down to any other Function below.

A flag in Pipeline view indicates that a Function is `Final`.



The Eval and first Drop Functions bearing the Final flag

Functions and Shared-Nothing Architecture

Cribl Edge is built on a shared-nothing architecture, where each Node and its **Worker Processes operate separately, and process events independently of each other**. This means that all Functions operate strictly in a Worker Process context – state is not shared across processes.

For example, consider two events that meet a Pipeline's criteria to be aggregated. If these two events arrive on **separate** Workers or Worker Processes, Cribl Edge will **not** aggregate them together.

This is particularly important to understand for certain Functions that might imply state-sharing, such as Aggregations, Rollup Metrics, Dynamic Sampling, Sampling, Suppress, etc.

If you have a large number of Worker Processes, consider implementing a distributed caching tier, such as Redis, to aggregate events across Workers. (See our Redis Function topic.)

Out-of-the-Box Functions

Cribl Edge ships with several Functions out-of-the-box, and you can chain them together to meet your requirements. For more details, see individual **Functions**, and the **Use Cases** section, within this documentation.

Accessing Event Fields with `__e`

The special variable `__e` represents the `(context)` event inside a JavaScript expression. Using `__e` with square bracket notation, you can access any field within the event object, for example, `__e['hostname']`. Functions use `__e` extensively. You also **must** use this notation for fields that contain a special character, like `-`, `.`, or `@`.

Supported JavaScript Standard

Cribl Edge supports the ECMAScript® 2015 Language Specification.

Very Large Integer Values

Cribl Edge's JavaScript implementation can safely represent integers only up to the Number.MAX_SAFE_INTEGER constant of about 9 quadrillion (precisely, {2^53}-1). Cribl Edge Functions will round down any integer larger than this, in Data Preview and other contexts. Trailing 0's might indicate such rounding down of large integers.

What Functions to Use When

- Add, remove, update fields: Eval, Lookup, Regex Extract
- Find & Replace, including basic `sed`-like, obfuscate, redact, hash, etc.: Mask, Eval
- Add GeoIP information to events: GeoIP
- Extract fields: Regex Extract, Parser
- Extract timestamps: Auto Timestamp
- Drop events: Drop, Regex Filter, Sampling, Suppress, Dynamic Sampling
- Sample events (e.g, high-volume, low-value data): Sampling, Dynamic Sampling
- Suppress events (e.g, duplicates, etc.): Suppress
- Serialize events to CEF format (send to various SIEMs): CEF Serializer
- Serialize / change format (e.g., convert JSON to CSV): Serialize
- Convert JSON arrays into their own events: JSON Unroll, XML Unroll
- Flatten nested structures (e.g., nested JSON): Flatten

- Aggregate events in real-time (i.e., statistical aggregations): Aggregations
- Convert events to metrics format: Publish Metrics, Prometheus Publisher (beta)
- Transforms dimensional metrics events into the OTLP (OpenTelemetry Protocol) format: OTLP Metrics
- Resolve hostname from IP address: Reverse DNS (beta)
- Extract numeric values from event fields, converting them to type `number`: Numerify
- Send events out to a command or a local file, via `stdin`, from any point in a Pipeline: Tee
- Convert an XML event's elements into individual events: XML Unroll
- Duplicate events in the same Pipeline, with optional added fields: Clone
- Break events **within**, instead of before they reach, a Pipeline: Event Breaker
- Add a text comment within a Pipeline's UI, to label steps without changing event data: Comment

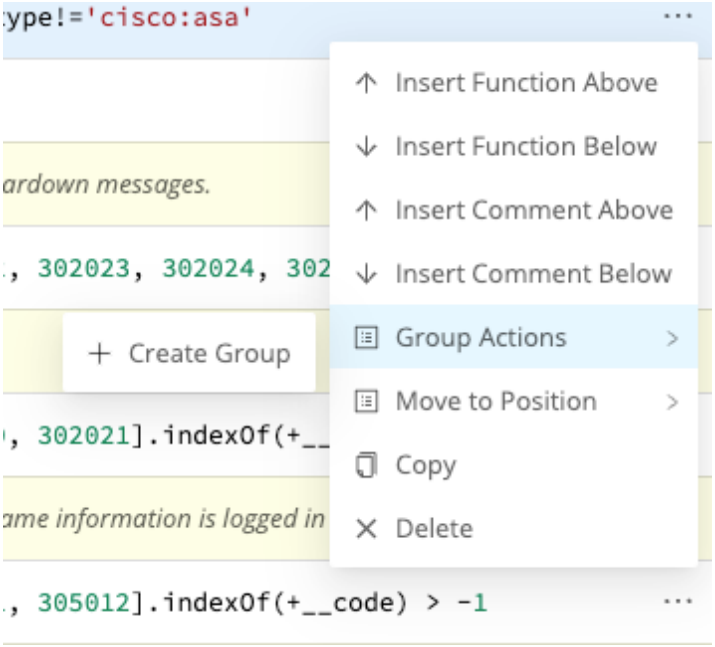
For more usage examples, download Cribl's Stream Cheat Sheet (Quick Reference Card).

Function Groups

A Function group is a collection of consecutive Functions that can be moved up and down a Pipeline's Functions stack together. Groups help you manage long stacks of Functions by streamlining their display. They are a UI visualization only: While Functions are in a group, those Functions maintain their global position order in the Pipeline.

Function groups work much like Route groups.

To build a group from any Function, click the Function's ••• (Options) menu, then select **Group Actions** > **Create Group**.



Creating a group

You'll need to enter a **Group Name** before you can save or resave the Pipeline. Optionally, enter a **Description**.

Once you've saved at least one group to a Pipeline, other Functions' ••• (Options) > **Group Actions** submenus will add options to **Move to Group** or **Ungroup/Ungroup All**.

You can also use a Function's left grab handle to drag and drop it into, or out of, a group. A saved group that's empty displays a dashed target into which you can drag and drop Functions.

Drag-and-drop target

14.1. Auto Timestamp

The Auto Timestamp Function extracts time to a destination field, given a source field in the event. By default, Auto Timestamp makes a first best effort and populates `_time`. When you add a sample (via paste or a local file), you should accomplish time and event breaking at the same time you add the data.

This Function allows fine-grained and powerful transformations to populate new time fields, or to edit existing time fields. You can use the Function's Additional timestamps section to create custom time fields using regex and custom JavaScript `strptime` functions.

The Auto Timestamp Function uses the same basic algorithm as the Event Breaker Function and the C.Time.timestampFinder() native method.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. The default `true` setting passes all events through the Function.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Source field: Field to search for a timestamp. Defaults to `_raw`.

Destination field: Field to place extracted timestamp in. Defaults to `_time`. Supports nested addressing.

Default timezone: Select a timezone to assign to timestamps that lack timezone info. Defaults to `Local`. (This drop-down includes support for legacy names: `EST5EDT`, `CST6CDT`, `MST7MDT`, and `PST8PDT`.)

Additional timestamps: To extract additional timestamp formats, click **Add Timestamp** to define each format. Each row will provide these fields:

- **Regex**: Regex, with first capturing group matching the timestamp.
- **Strptime format**: Select or enter the `strptime` format for the captured timestamp.

Advanced Settings

Time expression: Expression with which to format extracted time. Current time, as a JavaScript Date object, is in global `time`. Defaults to `time.getTime() / 1000`. You can access other fields' values via `__e.`

`<fieldName>`.

For details about Cribl Edge's Library (native) time methods, see: C.Time – Time Functions.

Start scan offset: How far into the string to look for a time string.

Max timestamp scan depth: Maximum string length at which to look for a timestamp.

Default time: How to set the time field if no timestamp is found. Defaults to **Current time**.

Two fields enable you to constrain (clamp) the parsed timestamp, to prevent the Function from mistakenly extracting non-time values as unrealistic timestamps:

- **Earliest timestamp allowed**: Enter a string that specifies the latest allowable timestamp, relative to now. (Sample value: `-42years`. Default value: `-420weeks`.) Parsed values earlier than this date will be set to the **Default time**.
- **Future timestamp allowed**: Enter a string that specifies the latest allowable timestamp, relative to now. (Sample value: `+42days`. Default value: `+1week`.) Parsed values after this date will be set to the **Default time**.

Format Reference

This references https://github.com/d3/d3-time-format#locale_format. Directives annotated with a `(†)` symbol might be affected by the locale definition.

```
%a - abbreviated weekday name. (†)
%A - full weekday name. (†)
%b - abbreviated month name. (†)
%B - full month name. (†)
%c - the locale's date and time, such as %x, %X. (†)
%d - zero-padded day of the month as a decimal number [01,31].
%e - space-padded day of the month as a decimal number [ 1,31]; equivalent to %_d.
%f - microseconds as a decimal number [000000, 999999].
%H - hour (24-hour clock) as a decimal number [00,23].
%I - hour (12-hour clock) as a decimal number [01,12].
%j - day of the year as a decimal number [001,366].
%m - month as a decimal number [01,12].
%M - minute as a decimal number [00,59].
%L - milliseconds as a decimal number [000, 999].
%p - either AM or PM. (†)
```

```
%Q - milliseconds since UNIX epoch.
%s - seconds since UNIX epoch.
%S - second as a decimal number [00,61].
%u - Monday-based (ISO 8601) weekday as a decimal number [1,7].
%U - Sunday-based week of the year as a decimal number [00,53].
%V - ISO 8601 week of the year as a decimal number [01, 53].
%w - Sunday-based weekday as a decimal number [0,6].
%W - Monday-based week of the year as a decimal number [00,53].
%x - the locale's date, such as %-m/%-d/%Y. (†)
%X - the locale's time, such as %-I:%M:%S %p. (†)
%y - year without century as a decimal number [00,99].
%Y - year with century as a decimal number.
%Z - time zone offset, such as -0700, -07:00, -07, or Z.
%% - a literal percent sign (%).
```

Complying with the Format

In order to use auto timestamping upon ingestion, the formatting used must match the `%Z` parameters above. E.g., this Function will automatically parse all of these formats:

- `2020/06/10T17:17:35.004-0700`
- `2020/06/10T17:17:35.004-07:00`
- `2020/06/10T17:17:35.004-07`
- `2020/06/10T10:17:35.004Z`
- `2020/06/10T11:17:35.004 EST`

To parse other formats, you can use the Additional Timestamps section's internal **Regex** or **Strptime Format** operators.

Basic Example

Filter: `name.startsWith('kumquats') && value=='specific string here'`

This will allow the Auto Timestamp Function to act only on events matching the specified parameters.

Sample event:

```
Sep 20 12:03:55 PA-VM 1,2019/09/20 13:03:58,CRIBL,TRAFFIC,end,2049,2019/09/20
14:03:58,314.817.108.226,10.0.0.102,314.817.108.226,10.0.2.65,cribl,,,incomplete,vsys1
forwarding-default,2018/09/20
```

```
13:03:58,574326,1,53722,8088,53722,8088,0x400064,tcp,allow,296,296,0,4,2018/09/20
13:03:45,7,any,0,730277,0x0,United States,10.0.0.0-10.255.255.255,0,4,0,aged-
out,0,0,0,0,,PA-VM,from-policy,,,0,,0,,N/A,0,0,0,0
```

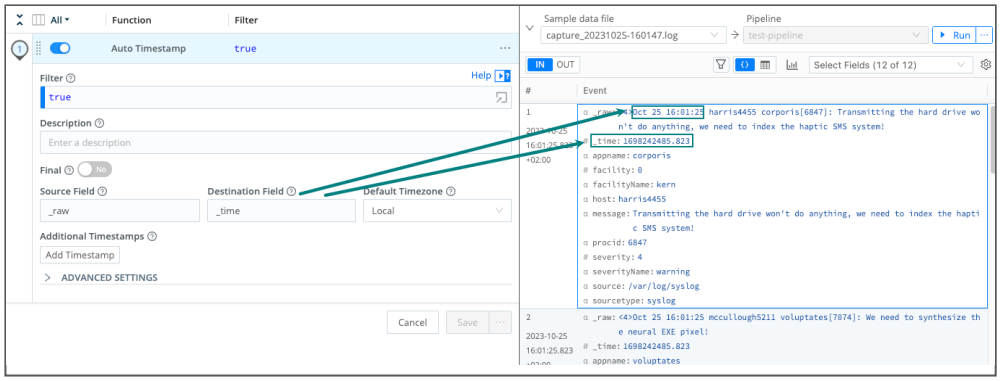
To add this sample (after creating an Auto Timestamp Function with the above **Filter** expression): Go to **Preview** > **Add a Sample** > **Paste a Sample**, and add the data snippet above. Do not make any changes to timestamping or line breaking, and select **Save as Sample File**.

By default, Cribl Edge will inspect the first 150 characters, and will extract the first valid timestamp it sees. You can modify this character limit under **Advanced Settings** > **Max Timestamp Scan Depth**.

Cribl Edge will grab the first part of the event, and will settle on the first matching value to display for `time`:

- `_time 1698242485.823`
- **GMT**: Oct 25 14:01:25 GMT
- **Your Local Time**: Oct 25 16:01:25 GMT +02:00

Because no explicit timezone has been set (under **Default Timezone**), `_time` will inherit the **Local** timezone, which in this example is `GMT +02:00`.



Timezone Dependencies and Details

Cribl Edge uses ICU for timezone information. It does not query external files or the operating system. The bundled ICU is updated periodically.

For additional timezone details, see: https://www.iana.org/time-zones.

Advanced Settings Example

The `datetime.strptime()` method creates a datetime object from the string passed in by the **Regex** field.

Here, we'll use `datetime.strptime()` to match a timestamp in AM/PM format at the end of a line.

Sample:

```
This is a sample event that will push the datetime values further on inside the event. This is still a sample event and finally here is the datetime information!: Server_UTC_Timestamp="04/27/2020 2:30:15 PM"
```

Max timestamp scan depth: `210`

Click to add **Additional timestamps**:

Regex: `(\d{1,2}\/\d{2}\/\d{4}\s\d{1,2}:\d{2}:\d{2}\s\w{2})`

Strptime format: `'%m/%d/%Y %H:%M:%S %p'`

Gnarly Details

- This Function supports the `%f` (microseconds) directive.
- Cribl Edge will truncate timestamps to three-digit (milliseconds) resolution, omitting trailing zeros.
- For further examples, see Extracting Timestamps from Messy Logs.

14.2. Aggregations

The Aggregations Function performs aggregate statistics on event data.

Each Worker Process executes this Function independently on its share of events. For details, see Functions and Shared-Nothing Architecture.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Time window: The time span of the tumbling window for aggregating events. Must be a valid time string (e.g., `10s`). Must match pattern `\d+[sm]$`.

Aggregates: Aggregation function(s) to perform on events.
E.g., `sum(bytes).where(action=='REJECT').as(TotalBytes)`. Expression format: `aggFunction(<FieldExpression>).where(<FilterExpression>).as(<outputField>)`. See more examples below.

When used without `as()`, the aggregate's output will be placed in a field labeled `<fieldName>_<aggFunction>`. If there are conflicts, the last aggregate wins. For example, given two aggregates – `sum(bytes).where(action=='REJECT')` and `sum(bytes)` – the latter one (`bytes_sum`) is the winner.

Group by Fields: Fields to group aggregates by. Supports wildcard expressions.

Evaluate fields: Set of key-value pairs to evaluate and add/set. Fields are added in the context of an aggregated event, before they're sent out. Does not apply to passthrough events.

Time Window Settings

Cumulative aggregations: If enabled, aggregations will be retained for cumulative aggregations when flushing out an aggregation table event. When set to `No` (the default), aggregations will be reset to `0` on

flush.

Lag tolerance: The lag tolerance represents the tumbling window tolerance to late events. Must be a valid time string (e.g., `10s`). Must match pattern `\d+[sm]$`.

Idle bucket time limit: The amount of time to wait before flushing a bucket that has not received events. Must be a valid time string (e.g., `10s`). Must match pattern `\d+[sm]$`.

Output Settings

Passthrough mode: Determines whether to pass through the original events along with the aggregation events. Defaults to `No`.

Metrics mode: Determines whether to output aggregates as metrics. Defaults to `No`, causing aggregates to be output as events.

Sufficient stats mode: Determines whether to output *only* statistics sufficient for the supplied aggregations. Defaults to `No`, meaning output richer statistics.

Output prefix: A prefix that is prepended to all of the fields output by this Aggregations Function.

Advanced Settings

Aggregation event limit: The maximum number of events to include in any given aggregation event. Defaults to unlimited. Must be at least `1`.

Aggregation memory limit: The memory usage limit to impose upon aggregations. Defaults to unlimited (i.e., the amount of memory available in the system). Accepts numerals with multiple-byte units, like KB, MB, GB, etc. (such: as `4GB`.)

Flush on stream close: If set to `Yes` (the default), aggregations will flush when an input stream is closed. If set to `No`, the Time Window Settings will control flush behavior; this can be preferable in cases like the following:

- Your input data consists of many small files.
- You are sending data to Prometheus. Enabling **Flush on stream close** can send Prometheus multiple aggregations from the same Worker Process for the same time period. Prometheus cannot tell the multiple aggregations apart, and will ingest only the first one.

Aggregation Functions

- `avg(expr:FieldExpression)`: Returns the average `expr` values.

- `count(expr:FieldExpression)`: If `expr` is omitted, returns the number of events received over the time window. If `expr` is given, returns the number of events seen where `expr` evaluates to a value other than null or undefined. For example, with these three events in the same time window:
  - `{ _time: 20, _val: 5 }`
  - `{ _time: 22 }`
  - `{ _time: 24, _val: null }`

  `count()` would return `3`, while `count(_val)` would return `1`.
- `dc(expr: FieldExpression, errorRate: number = 0.01)`: Returns the estimated number of distinct values of `expr`, within a relative error rate. Lower error rates increase the accuracy of the result, at the cost of using more memory. For example, with these three events in the same time window:
  - `{ _time: 20, _name: 'Alice' }`
  - `{ _time: 22, _name: 'Bob' }`
  - `{ _time: 24, _name: 'Alice' }`

  `dc(_name)` would return `2`.
- `distinct_count(expr: FieldExpression, errorRate: number = 0.01)`: Identical to `dc(expr)`.
- `earliest(expr:FieldExpression)`: Returns the earliest (based on `_time`) observed `expr` value.
- `first(expr:FieldExpression)`: Returns the first observed `expr` value.
- `histogram(expr:FieldExpression, buckets: number[])`. Returns the average of the values of `expr` and generates a field with the same name as the aggregated output field, suffixed with `_data`.
  - `buckets` must contain at least one numeric value.

  The `_data` field in the output contains three pieces of information:
  - `_sum` – the sum of all values of `expr`.
  - `_count` – the number of events seen where `expr` evaluates to a value other than null or undefined.
  - `_buckets` – an object whose keys are the buckets defined in the function, and whose values are the number of values that fall in that histogram bucket. In addition to the buckets defined by the user, the object will include a bucket labeled `Infinity`, into which all values will be placed. A value of `x` falls into a histogram bucket if it is less than or equal to the value of the bucket. A value of `45` would fall into a bucket with value `50`, but not one with value `40`. Counts are cumulative; all values counted in a bucket will also be counted in every bucket larger than it. For example, with these three events in the same time window:
    - `{ _time: 20, age: 20 }`
    - `{ _time: 22, age: 25 }`
    - `{ _time: 24, age: 30 }`

`histogram(age, [10, 25, 40])` would result in the following on the output event:

```
{
  age_histogram: 25
  age_histogram_data: {
    _count: 3,
    _sum: 75
    _buckets: {
      10: 0,
      25: 2,
      40: 3,
      Infinity: 3
    }
  }
}
```

- `last(expr:FieldExpression)`: Returns the last observed `expr` value.
- `latest(expr:FieldExpression)`: Returns the latest (based on `_time`) observed `expr` value.
- `list(expr:FieldExpression[, max:number, excludeNulls: boolean = true])`: Returns a list of the observed values of `expr`.
    - Optional `max` parameter limits the number of values returned. If omitted, the default is `100`. If set to `0`, will return all values.
    - Optional `excludeNulls` boolean excludes null and undefined values from results. If included, defaults to `true`.
- `max(expr:FieldExpression)`: Returns the maximum `expr` value.
- `median(expr:FieldExpression)`: Returns the middle value of the sorted parameter.
- `min(expr:FieldExpression)`: Returns the minimum `expr` value.
- `mode(expr: FieldExpression[, excludeNulls: boolean = true])`: Returns the single most frequently encountered `expr` value.
    - Optional `excludeNulls` boolean excludes null and undefined values from results. If included, defaults to `true`.
- `per_second(expr:FieldExpression)`: Returns the rate of change of the values of `expr` over the aggregate time window. This is equivalent to `rate(expr, '1s')`. For example, with these three events in the same time window:
    - `{ _time: 20, _val: 5 }`
    - `{ _time: 22, _val: 15 }`
    - `{ _time: 24, _val: 35 }`

`per_second(_val)` would give back `(35 - 5) / (24 - 20) = 7.5`, meaning that `_val` increased by 7.5 every second over the time window.

- `perc(level: number, expr: FieldExpression)`: Returns `<level>` percentile value of the numeric `expr` values.
- `rate(expr:FieldExpression, timeString: string = '1s')`: Returns the rate of change of the values of `expr` over the aggregate time window. Calculated as (latest value - earliest value) / (latest timestamp - earliest timestamp) * number of seconds in `timeString`. For example, with these three events in the same time window:
  - `{ _time: 20, _val: 5 }`
  - `{ _time: 22, _val: 15 }`
  - `{ _time: 24, _val: 35 }`

  `rate(_val, '2s')` would give back `(35 - 5) / (24 - 20) * 2 = 15`, meaning that `_val` increased by 15 every 2 seconds over the time window.
- `stdev(expr:FieldExpression)`: Returns the sample standard deviation of the `expr` values.
- `stdevp(expr:FieldExpression)`: Returns the population standard deviation of the `expr` values.
- `sum(expr:FieldExpression)`: Returns the sum of the `expr` values.
- `summary(expr:FieldExpression)[, quantiles: number[]])`: Returns the average of the `expr` values and generates a field with the same name as the aggregate output field, suffixed with `_data`.
  - Optional: `quantiles` values must be between `0` and `1` inclusive.

  The `_data` field in the output contains three pieces of information:
  - `_sum` – the sum of all `expr` values.
  - `_count` – the number of events seen where `expr` evaluates to a value other than null or undefined.
  - `_quantiles` – an object whose keys are the quantiles defined in the function, and whose values are the quantile value across the `expr` values. For example, a quantile key `0.5` and value `500` would indicate that the 50th percentile (or median) of the values seen was 500.
  - For example, with these three events in the same time window:
    - `{ foo: 100 }`
    - `{ foo: 200 }`
    - `{ foo: 300 }`

    `summary(foo, [0.25, 0.5, 0.75])` would result in the following on the output event:

```
{
  foo_summary: 25
  foo_summary_data: {
    _count: 3,
    _sum: 75
    _quantiles: {
      0.25: 100,
      0.5: 150,
      0.75: 225
    }
  }
}
```

- `sumsq(expr:FieldExpression)`: Returns the sum of squares of the `expr` values.
- `top(expr: FieldExpression, count: number[, excludeNulls: boolean = true])`: Returns the most frequently encountered `expr` values, up to `count` number of results.
  - `count` parameter must be a positive integer.
  - Optional `excludeNulls` boolean excludes null and undefined values from results. If included, defaults to `true`.
- `values(expr:FieldExpression[, max:number, errorRate:number, excludeNulls: boolean = true])`: Returns a list of distinct `expr` values.
  - Optional `max` parameter limits the number of values returned; if omitted, the default is `0`, meaning return all distinct values.
  - Optional `errorRate` parameter controls how accurately the function counts "distinct" values. Range is `0-1`; if omitted, the default value is `0.01`. Higher values allow higher error rates (fewer unique values recognized), with the offsetting benefit of less memory usage.
  - Optional `excludeNulls` boolean excludes null and undefined values from results. If included, defaults to `true`.
- `variance(expr:FieldExpression)`: Returns the sample variance of the `expr` values.
- `variancep(expr:FieldExpression)`: Returns the population variance of the `expr` values.

Safeguarding Data

Upon shutdown, Cribl Edge will attempt to flush the buffers that hold aggregated data, to avoid data loss. If you set a **Time window** greater than 1 hour, Cribl recommends adjusting the **Aggregation memory limit** and/or **Aggregation event limit** to prevent the system from running out of memory.

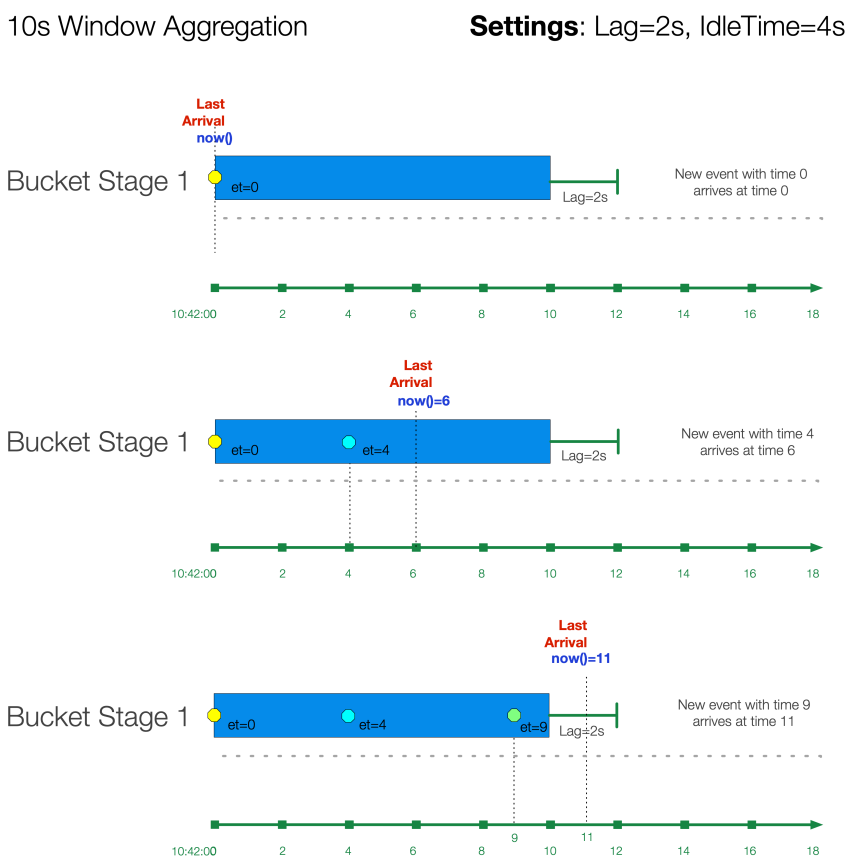
This is especially necessary for high-cardinality data. (Both settings default to unlimited, but we recommend setting defined limits based on testing.)

How Do Time Window Settings Work?

Lag Tolerance

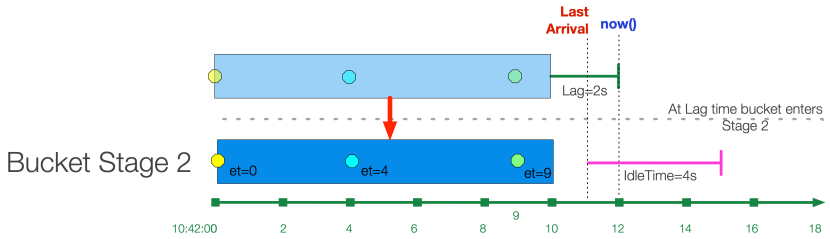
As events are aggregated into windows, there is a good chance that most will arrive later than their event time. For instance, given a `10s` window (`10:42:00 - 10:42:10`), an event with timestamp `10:42:03` might come in 2 seconds later at `10:42:05`.

In several cases, there will also be late, or lagging, events that will arrive **after** the latest time window boundary. For example, an event with timestamp `10:42:04` might arrive at `10:42:12`. Lag Tolerance is the setting that governs how long to wait – after the latest window boundary – and still accept late events.



The "bucket" of events is said to be in Stage 1, where it's still accepting new events, but it's not yet finalized. Notice how in the third case, an event with event time `10:42:09` arrives 1 second past the window boundary at `10:42:11`, but it's still accepted because it happens before the lag time expires.

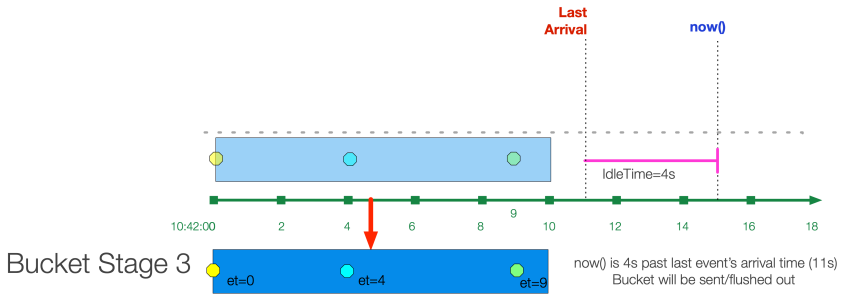
After the lag time expires, the bucket moves to Stage 2.



If the bucket is created from a historic stream, then the bucket is initiated in Stage 2. Lag time is not considered. A "historic" stream is one where the latest time of a bucket is before `now()`. E.g., if the window size is 10s, and `now()=10:42:42`, an event with `event_time=10` will be placed in a Stage 2 bucket with range `10:42:10 - 10:42:20`.

Idle Bucket Time Limit

While Lag Tolerance works with event time, Idle Bucket Time Limit works on arrival time (i.e., real time). It is defined as the amount of time to wait before flushing a bucket that has not received events.



After the Idle Time limit is reached, the bucket is "flushed" and sent out of the system.

Examples

Assume we're working with VPC Flowlog events that have the following structure:

```
version account_id interface_id srcaddr dstaddr srcport dstport protocol packets
bytes start end action log_status
```

For example:

```
2 99999XXXXX eni-02f03c2880e4aaa3 10.0.1.70 10.0.1.11 9999 63030 6 6556 262256
1554562460 1554562475 ACCEPT OK 2 496698360409 eni-08e66c4525538d10b 37.23.15.38
10.0.2.232 4373 8108 6 1 52 1554562456 1554562466 REJECT OK
```

Scenario A:

Every 10s, compute sum of `bytes` and output it in a field called `TotalBytes`.

Time Window: `10s` Aggregations: `sum(bytes).as(TotalBytes)`

Scenario B:

Every 10s, compute sum of `bytes`, output it in a field called `TotalBytes`, group by `srcaddr`.

Time Window: `10s` Aggregations: `sum(bytes).as(TotalBytes)` Group by Fields: `srcaddr`

Scenario C:

Every 10s, compute sum of `bytes` but only where action is `REJECT`, output it in a field called `TotalBytes`, group by `srcaddr`.

Time Window: `10s` Aggregations: `sum(bytes).where(action=='REJECT').as(TotalBytes)` Group by Fields: `srcaddr`

Scenario D:

Every 10s, compute sum of `bytes` but only where action is `REJECT`, output it in a field called `TotalBytes`. Also, compute distinct count of `srcaddr`.

Time Window: `10s` Aggregations:

```
sum(bytes).where(action=='REJECT').as(TotalBytes)
distinct_count(srcaddr).where(action=='REJECT')
```

For further examples, see Engineering Deep Dive: Streaming Aggregations Part 2 – Memory Optimization.

14.3. CEF Serializer

The CEF Serializer takes a list of fields and/or values, and formats them in the Common Event Format (CEF) standard. CEF defines a syntax for log records. It is composed of a standard prefix, and a variable extension formatted as a series of key-value pairs.

Format

`CEF:Version|Device Vendor|Device Product|Device Version|Device Event Class ID|Name|Severity|[Extension]`

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Output field: The field to which the CEF formatted event will be output. Nested addressing supported. Defaults to `_raw`.

Header Fields

CEF Header field definitions. The field values below will be written pipe (|)–delimited in the Output Field. Names cannot be changed. Values can be computed with JS expression, or can be constants.

- **cef_version**: Defaults to `CEF:0`.
- **device_vendor**: Defaults to `Cribl`.
- **device_product**: Defaults to `Cribl`.
- **device_version**: Defaults to `C.version`.
- **device_event_class_id**: Defaults to `420`.
- **name**: Defaults to `Cribl Event`.
- **severity**: Defaults to `6`.

Extension Fields

CEF Extension field definitions. Field names and values will be written in `key=value` format. Select each field's Name from the drop-down list. Values can be computed with JS expressions, or can be constants.

Example

For each CEF field, allowed values include strings, plus any custom Cribl function. For example, if using a lookup:

Name: `Name` Value expression: `C.Lookup('lookup-exact.csv', 'foo').match('abc', 'bar')`

This can be used for any of the `CEF` **Header Fields**.

The resulting event has the following structure for an **Output Field** set to `_CEF_out`:

```
_CEF_out:CEF:0|Cribl|Cribl|42.0-61c12259|420|Business Group
6|6|c6a1Label=Colorado_Ext_Bldg7
```

14.4. Chain

The Chain Function does one thing: It chains data processing from a Pipeline or Pack to another Pipeline or Pack. This can be useful for sequential processing, or just to separate groups of related Functions into discrete Pipeline or Pack units that make intuitive sense.

Control Flow

The chained Pipeline or Pack will, upon completion, normally return control to the parent Pipeline/Pack containing the Chain Function. However, if the chained Pipeline/Pack contains any Function with the **Final** flag enabled, processing will stop there. In this case, back in the parent Pipeline/Pack, no Function below Chain will execute.

Cycle Detection and Throughput

The Chain Function includes guardrails against circular references. In v.4.3 and later, Cribl Edge detects cycles when you configure your Pipeline. This early detection (compared to earlier versions' runtime detection) means that:

- If you try to save a Pipeline with a cyclical reference, Cribl Edge will throw an error.
- If an existing Pipeline's configuration contains a cycle, Cribl Edge will disable the final Chain Function involved in the cycle.

Despite these safeguards, Cribl recommends that you keep chained configurations simple and understandable by all your users. Also, keep in mind that inserting a Chain Function can impose a slight performance hit, compared to including all processing in the original Pipeline or Pack.

Pipeline Versus Pack Scope

You will see different scope restrictions when using Chain in a Pipeline versus in a Pack:

- In a Pipeline, the **Processor** drop-down displays both Pipelines and Packs as targets to chain to.
- In a Pack, the **Processor** drop-down offers only Pipelines contained within that Pack.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

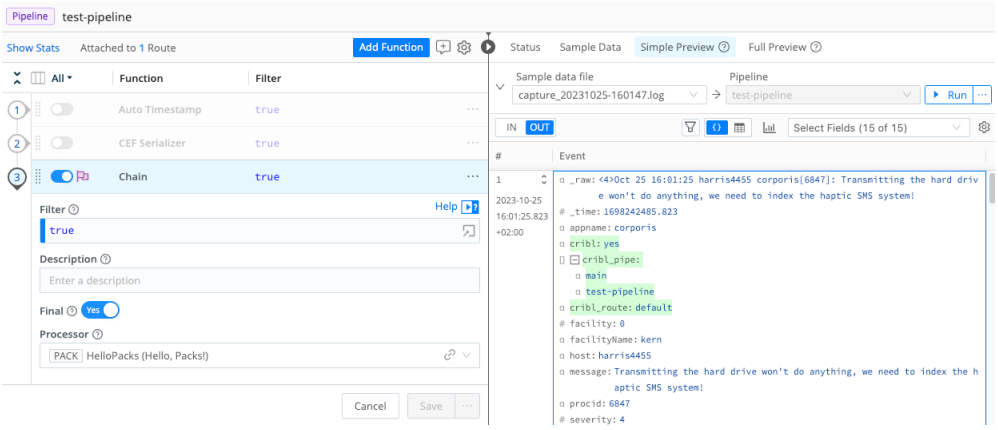
Description: Optionally, add a simple description of this Function's purpose. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`. (Note that this will not prevent data from flowing to the Function's defined **Processor**.)

Processor: Use this drop-down to select a configured Pipeline or Pack through which to forward events.

Example

This shows a simple preview of a `pipeline-1` Pipeline, which chains to a `pipeinpipe` Pipeline. Notice that each event's added `cribl_pipe` field lists all Pipelines/Packs through which the event was chained.



cribl_pipe field shows whole processing path

14.5. Clone

The Clone Function clones events, with optional added fields. Cloned events will be sent to the same Destination as the original event, because they are in the same Pipeline. If your Destination is an Output Router, you can use filters to send cloned events to different Destinations.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Clones: Create clones with the specified fields added and set.

Fields: Set of key-value pairs to add. Nested addressing is supported.

Examples

Staging Example

In this example, the Destination will receive a clone with an `env` field set to `staging`.

Field: `env` **Value**: `staging`

Index Example

In this scenario, we insert a Clone Function at the beginning of a Pipeline to create cloned events. We can later use these events as a baseline to compare against the original events, after various Functions have processed them.

You can assign any meaningful fields to the cloned events – anything that will help you identify them when comparing. This example simply assigns a key-value pair of `index: clones`.

Field: `index` **Value**: `clones`

To keep the cloned events from being processed by Functions later in the same Pipeline, you'll need to specify `index!='clones'` in their **Filter** expressions.

14.6. Code

If you need to operate on data in a way that can't be accomplished with Cribl Edge's out-of-the-box Functions, the Code Function enables you to encapsulate your own JavaScript code. This Function is available in Cribl Edge 3.1+, and imposes some restrictions for security reasons.

Restrictions

Generally speaking, anything forbidden in JavaScript strict mode is forbidden in the context of the Code Function. Specifically, the following are **not allowed**:

- `console`, `eval`, `uneval`, `Function (constructor)`, `Promises`, `setTimeout`, `setInterval`, `global`, `globalThis`, `window`, and `set`.

Code Functions **can** include `for` loops, `while` loops, and JavaScript methods such as `map`, `reduce`, `forEach`, `some`, and `every`. For further details, see Supported JavaScript Options.

Cribl Edge's predefined Functions, such as Eval, cover the vast majority of scenarios that users typically need to implement. You should use Code Functions only as a last resort, when you need to construct a complex block of code.

Also, only skilled JavaScript developers should define Code Functions. This is to avoid unintended results – such as creating infinite loops, or otherwise failing to return – that could needlessly add to your throughput burden.

Usage

When added to a Pipeline, the Code Function offers the following configuration options:

Filter: JavaScript filter expression that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Optionally, add a simple description of this Function.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Code: The mini-editor where you type your JavaScript code.

Advanced Settings

Maximum number of iterations: The maximum number of iterations per instance of this Code Function. Defaults to `5,000`; highest allowed value is `10000`.

Error log sample rate: Specifies the rate at which this Code Function logs errors. For example, a value of `1` logs every error; a value of `10` logs every tenth error. The highest allowed value is `5000`. Defaults to `1`.

Use unique log channel: When enabled, Cribl Edge sends logs from this function to a unique channel in the form `func:code:${pipelineName}:${functionIndex}`. Toggle off to use the generic `func:code` log channel instead.

Notes and Examples

Functions (including the Code Function) always use the special variable `__e` to access the `(context)` event inside JavaScript expressions.

Possibly the simplest Code Function creates a new field and then assigns it a value:

A Basic Code Function

```
__e['foo'] = 'Hello, Goats!'
```

For more ambitious implementations, see Code Function Examples.

JavaScript Support

Cribl Edge supports the ECMAScript® 2015 Language Specification.

With some exceptions, the Code Function supports the options described in the following MDN JavaScript Guide topics:

- Expressions and Operators
- Global variables
- Control flow and error handling
- Loops and iteration
- Functions
- Numbers and dates
- Text formatting
- Regular Expressions

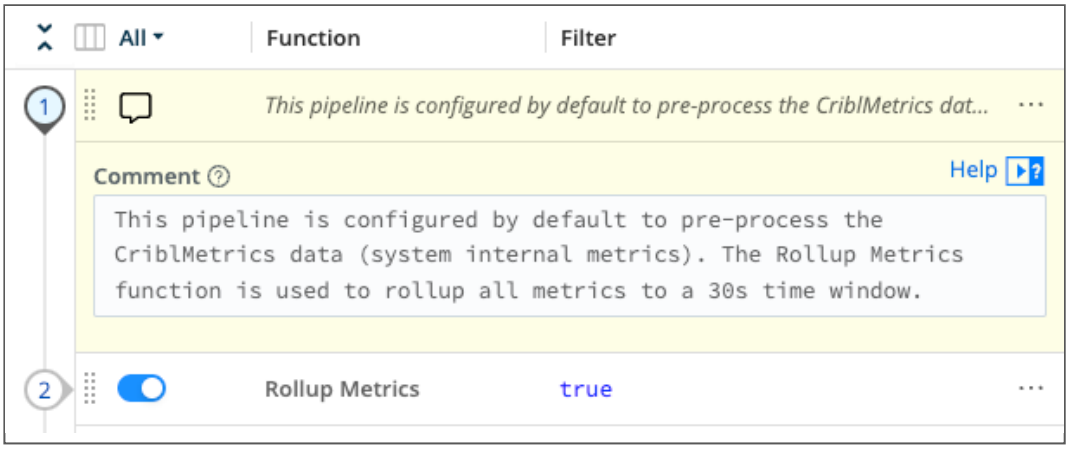
- Indexed collections
- Keyed collections
- Working with Objects

14.7. Comment

The Comment Function adds a text comment in a Pipeline. It makes no changes to event data. The added comment is visible only within the Pipeline UI, where it is useful for labeling Pipeline steps. The Comment Function always has a minimum, non-zero processing time.

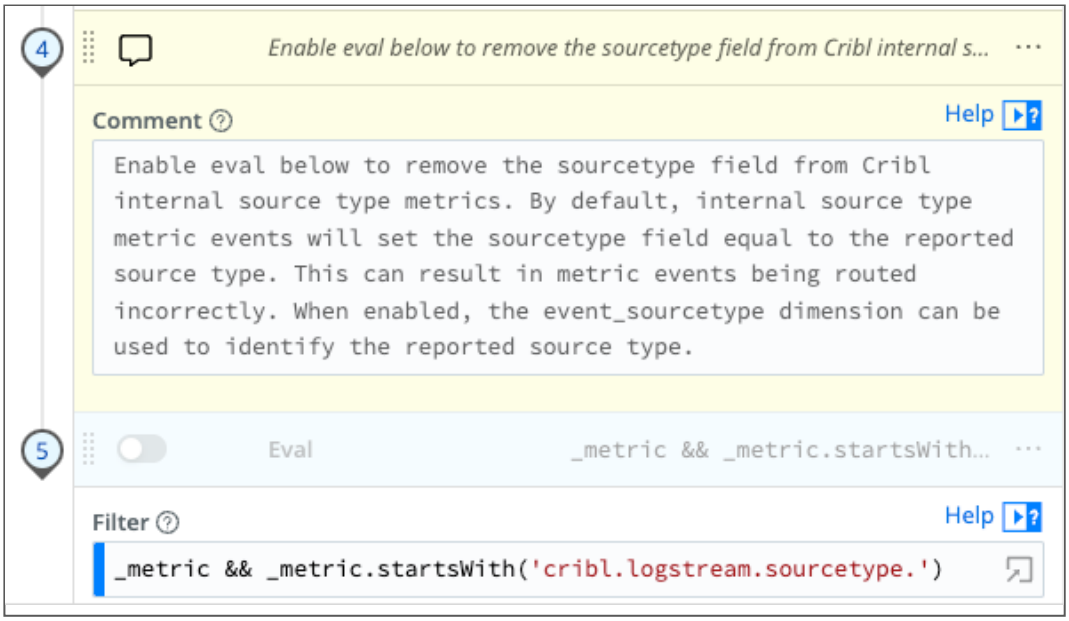
Usage

Comment: Add your comment as plain text in this field.



Examples

This comment labels the Pipeline's next function:



14.8. DNS Lookup

The DNS Lookup Function offers two operations useful in enriching security and other data:

- DNS lookups based on host name as text, resolving to A record (IP address) or to other record types.
- Reverse DNS Lookup. (This duplicates the behavior of Cribl Edge's prior Reverse DNS Function, which was deprecated as of v.2.4, and has been removed as of v.4.0.)

To reduce DNS lookups and minimize latency, the DNS Lookup Function incorporates a configurable DNS cache (including resolved and unresolved lookups). If you need additional caching, consider enabling OS-level DNS caching on each Cribl Edge Worker that will execute this Function. (OS-level caching options include DNSMasq, nscd, systemd-resolved, etc.)

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

DNS Lookup Fields Section

Lookup field name: Name of the field containing the domain to look up.

Resource record type: DNS record type (RR) to return. Defaults to `A`' record.

Output field name: Lookup result(s) will be added to this field. Leave blank to overwrite the original field specified in **Lookup field name**.

Reverse DNS Lookup Field(s) Section

Lookup field name: Name of the field containing the IP address to look up.

If the field value is not in IPv4 or IPv6 format, the lookup is skipped.

Output field name: Name of the field in which to add the resolved hostname. Leave blank to overwrite the original field specified in **Lookup field name**.

Fallback Resolving Methods

This section contains settings for resolving DNS short names.

Use /etc/resolv.conf: Toggle to `Yes` if you want Cribl Edge to resolve DNS short names using the search or domain directive from `/etc/resolv.conf`. Defaults to `No`. Not applicable for Windows hosts.

Use search or domain fallback(s): Add fallback value(s) for the DNS resolver to use when it cannot resolve a DNS short name.

Fall back to DNS.lookup(): Toggle to `Yes` if you want Cribl Edge to make a `DNS.lookup()` call to resolve a DNS short name it can't resolve by other methods.

If you enable all of these methods at once. Cribl Edge will attempt to resolve short names by trying each method in the following order:

1. **Use /etc/resolv.conf**.
2. **Use search or domain fallback(s)**.
3. **Fall back to DNS.lookup()**.

If Cribl Edge can't resolve the short name after trying all of these methods, it will drop the output field from the event.

Potential Performance Impact

Making a `DNS.lookup()` call might degrade performance in unrelated areas of Cribl Edge. Use this method only if you have not been able to resolve DNS short names using other settings.

Advanced Settings

DNS server(s) overrides: IP address(es), in RFC 5952 format, of the DNS server(s) to use for resolution. IPv4 examples: `1.1.1.1`, `4.2.2.2:53`. IPv6 examples: `[2001:4860:4860::8888]`, `[2001:4860:4860::8888]:1053`. If this field is not specified, Cribl Edge will use the system's DNS server.

Cache time to live (minutes): Determines the interval on which the DNS cache will expire, and its contents will be refetched. Defaults to `30` minutes. Use `0` to disable cache expiration/refresh behavior.

Maximum cache size: Maximum number of DNS resolutions to cache locally. Before changing the default `5000`, consider the implications for your system. Higher values will increase memory usage. Highest allowed value is `100000`.

Monitoring DNS Cache State

When the DNS Lookup Function succeeds in looking up a value it has stored in the DNS cache, that is called a **hit**. Each hit enables the Function to skip contacting a DNS server outside of Cribl Edge to perform the DNS lookup. The resulting performance improvement is the main benefit that the DNS cache provides.

When the DNS Lookup Function looks for a value in the DNS cache, but cannot find it, that is called a **miss**. In this case, the Function must contact a DNS server to perform the lookup.

Cribl Edge logs the DNS cache's state in the `func:dns_lookup` channel. The table below explains some typical messages.

Log Message	Meaning
`hits`	Number of successful lookups from cache.
`misses`	Number of unsuccessful lookups from cache.
`keys`	Number of entries in cache.
`ksize`	Approximate amount of memory occupied by keys, in bytes.
`vsize`	Approximate amount of memory occupied by values, in bytes.

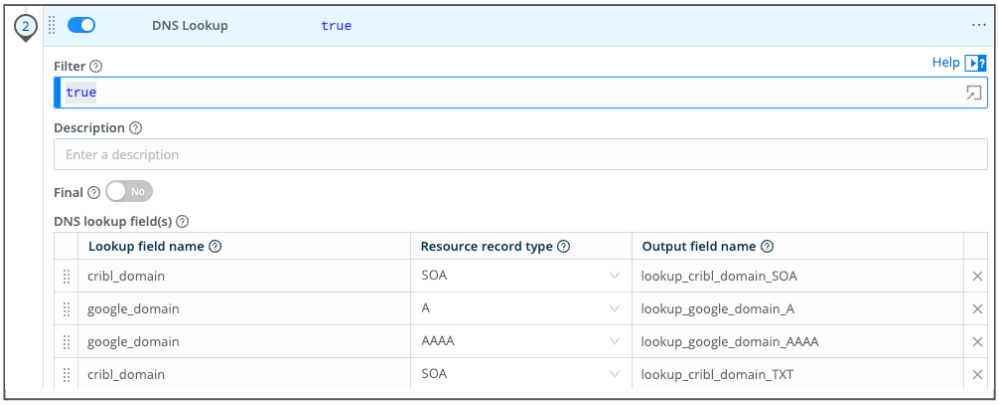
See Monitoring for a general explanation of logging in Cribl Edge.

Example

This example Pipeline chains two Functions. First, we have an **Eval** Function that defines key-value pairs for two alphabetical domain names and two numeric IP addresses.

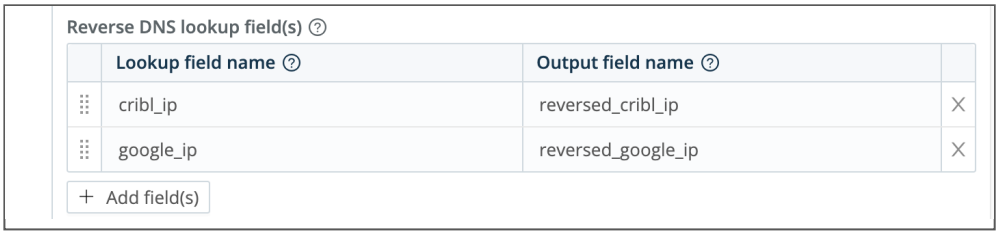
DNS Lookup: Eval Function

Next, the DNS Lookup Function looks up several record types for the two domain names, placing each retrieved record type in its own output field.



DNS Lookup: multiple record types

Finally, the same Function's **Reverse DNS lookup** section retrieves domain names for the two IP addresses.

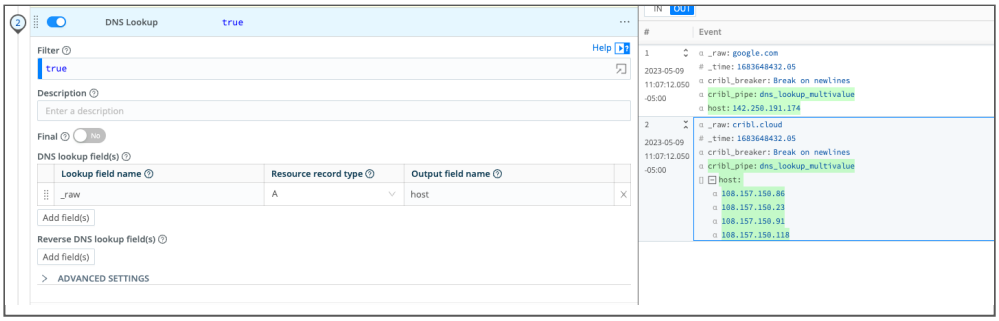


DNS Lookup: reverse lookups

Working with Multi-value Results

DNS records can contain single values mixed together with multi-value results. While this can be challenging to work with, the right code can adjust your results to the desired values.

In the screenshot below, the DNS lookup for `google.com` (event 1) returns a single record, while the `cribl.cloud` lookup (event 2) returns four addresses.



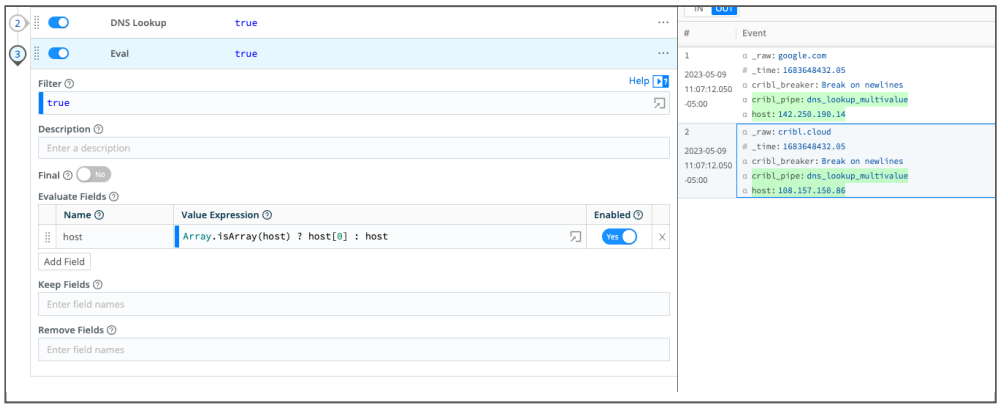
DNS lookup with multi-value result

Sometimes a receiver is unable to consume the data in an array format. One solution to this problem is to pick the first result and convert it into a string. Here's a JavaScript eval to do this:

```
Array.isArray(host) ? host[0] : host;
```

This function checks whether the value of the `host` field is an array.

- If true, the result is the left side of the ternary expression – in this case, `host[0]`, which is the first element in the array, converted into a string. See event 2 in the screenshot below.
- If false, the result is the single-value string in the `host` field. See event 1.



DNS lookup multi-value result converted into a single value

Another option is to **randomly** pick an address from the result. To do this, use a modified version of the JavaScript eval:

```
Array.isArray(host) ? host[Math.floor(Math.random() * host.length)] : host
```

Here, instead of picking the first element, the code uses `Math.floor(Math.random() * host.length)` to randomly pick an element from the array.

14.9. Drop

The Drop Function drops (deletes) any events that meet its Filter expression. This is useful when you want to prevent certain events from continuing to a Pipeline's downstream Functions.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Examples

Scenario A:

Assume that we care only about errors, so we want to filter out any events that contain the word "success," regardless of case: "success," "SUCCESS," etc.

In our Drop Function, we'll use the JavaScript `search()` method to search the `_raw` field's contents for our target pattern. We know that `search()` returns a non-negative integer to indicate the starting position of the first match in the string, or `-1` if no match. So we can evaluate the Function as true when the return value is `>= 0`.

Filter: `_raw.search(/success/i)>=0`

Scenario B:

You can filter out specific JSON events based on their key-value pairs.

The following Filter expression uses a strict inequality operator to check, per event, whether `channel` has a different data type **or** value from `auth`. Matching events will drop, ensuring that only events with `"channel":"auth"` will pass to the next Function.

Filter: `channel !== 'auth'`

14.10. Dynamic Sampling

The Dynamic Sampling Function filters out events based on an expression, a sample mode, and the volume of events. Your sample mode's configuration determines what percentage of incoming events will be passed along to the next step.

Each Worker Process executes this Function independently on its share of events. For details, see Functions and Shared-Nothing Architecture.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Sample mode: Defines how sample rate will be derived. For formulas and usage details, see Sample Modes below. Supported methods:

- Logarithmic (the default): `log(previousPeriodCount)`.
- Square root: `sqrt(previousPeriodCount)`.

Sample group key: Expression used to derive sample group key. For example: `${domain}:${httpCode}`. Each sample group will have its own derived sampling rate, based on the volume of events. Defaults to `` `${host}` ``.

All events without a host field passing through the Function will be associated with the same group and sampled the same.

Advanced Settings

- **Sample period Sec**: How often (in seconds) sample rates will be adjusted. Defaults to `30`.
- **Minimum events**: Minimum number of events that must be received, in previous sample period, for sampling mode to be applied to current period. If the number of events received for a sample group is less than this minimum, a sample rate of 1:1 is used. Defaults to `30`.
- **Max sampling rate**. Maximum sampling rate. If the computed sampling rate is above this value, the rate will be limited to this value.

How Does Dynamic Sampling Work

Compared to static sampling, where users must first select a sample rate, Dynamic Sampling allows for **automatically adjusting** sampling rates, based on the volume of incoming events per sample group. This Function allows users to set only the aggressiveness/coarseness of this adjustment. Square Root is more aggressive than Logarithmic mode.

As an event passes through the Function, it's evaluated against the Sample Group Key expression to determine the sample group it will be associated with. For example, given an event with these fields: `...ip=1.2.3.42, port=1234...`, and a Sample Group Key of `` `${ip}:${port}` ``, the event will be associated with the `1.2.3.42:1234` sample group.

If the Sample Group Key is left at its `` `${host}` `` default, all events without a host will be associated with the same group and sampled the same.

When a sample group is new, it will initially have a sample rate of `1:1` for `Sample Period` seconds (this value defaults to 30 seconds). Once `Sample Period` seconds have elapsed, a sample rate will be derived based on the configured `Sample Mode`, using the sample group's event volume during the **previous** sample period.

For example, assuming a Logarithmic Sample Mode:

**Period 0 (first 30s): ** Number of events in sample group: `1000`, Sample Rate: `1:1`, Events allowed: `ALL`
Sample Rate calculation for **next** period: `Math.ceil(Math.log(1000)) = 7`

Period 1 (next 30s) – Number of events in sample group: `4000`, Sample Rate: `7:1`: Events allowed: `572`
Sample Rate calculation for **next** period: `Math.ceil(Math.log(4000)) = 9`

Period 2 (next 30s) – Number of events in sample group: `12000`, Sample Rate: `9:1`: Events allowed: `1334`
Sample Rate calculation for **next** period: `Math.ceil(Math.log(12000)) = 10`

Period 3 (next 30s) – Number of events in sample group: `2000`, Sample Rate: `10:1`: Events allowed: `200`
Sample Rate calculation for **next** period: `Math.ceil(Math.log(2000)) = 8`
...

Sample Modes

1. Logarithmic – The sample rate is derived, for each sample group, using a natural log: `Math.ceil(Math.log(lastPeriodVolume))`. This mode is **less aggressive**, and drops fewer events.

2. Square Root – The sample rate is derived, for each sample group, using: `Math.ceil(Math.sqrt(lastPeriodVolume))`. This mode is **more aggressive**, and drops more events.

Example

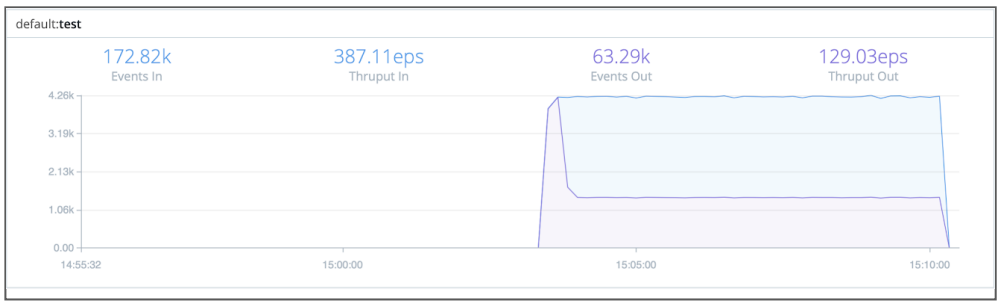
Here's an example that illustrates the effectiveness of using the Square Root sample mode.

Settings:

Sample Mode: `Square Root` Sample Period (sec): `20` Minimum Events: `3` Max. Sampling Rate: `3`

Results:

Events In: 4.23K Events Out: 1.41K



In this generic example, we reduced the incoming event volume from 4.23K to 1.41K. Your own results will vary depending on multiple parameters – the **Sample Group Key**, **Sample Period**, **Minimum Events**, **Max Sampling Rate**, and rate of incoming events.

For further examples, see Getting Smart and Practical With Dynamic Sampling.

14.11. Eval

The Eval Function adds or removes fields from events. (In Splunk, these are index-time fields.)

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Evaluate fields: Table of event fields to evaluate and add/set. Click **Add Field** to add each field as a key-value pair, in a row with the following controls:

- **Name**: Enter the new (or modified) field's name.
- **Value Expression**: Enter a JavaScript expression to set the field's value – this can be a constant. Expressions can contain nested addressing. When you insert JavaScript template literals, strings intended to be used as values must be delimited with single quotes, double quotes, or backticks. (For details, see Cribl Expression Syntax.)
- **Enabled**: Toggle this off to disable evaluating individual expressions, while retaining their configuration in the table. Useful for iterative development and debugging.

Keep fields: List of fields to keep in the event after processing by this Function. Supports wildcards (*) and nested addressing. Takes precedence over **Remove fields** (below). To reference a parent object and all children, you must use the (*) wildcard. For example, if `_raw` is converted to an object then use `_raw*` to refer to itself and all children.

Remove fields: List of fields to remove from the event. Supports wildcards (*) and nested addressing. Cannot remove fields that match against the **Keep fields** list. Cribl Edge internal fields that start with `__` (double underscore) **cannot** be removed via wildcard. Instead, you must specify them individually. For example, `__myField` cannot be removed by specifying `__myF*`.

Using Keep and Remove

A field matching an entry in *both* **Keep** (wildcard or not) and **Remove** will *not* be removed. This is useful for implementing "remove all but" functionality. For example, to keep only `_time, _raw, source, sourcetype, host`, we can specify them all in **Keep**, while specifying `*` in **Remove**.

Negated terms are supported in both **Keep fields** and **Remove fields**. The list is order-sensitive when negated terms are used. Examples:

- `!foobar, foo*` means "All fields that start with 'foo' except `foobar`."
- `!foo*, *` means "All fields except for those that start with 'foo'."

Examples

Note that Functions use the special variable `__e` to access the `(context)` event inside JavaScript expressions.

Scenario A: Create field `myField` with static value of `value1`:

- **Name:** `myField`
- **Value Expression:** `'value1'`

Scenario B: Set field `action` to `blocked` if `login==error`:

- **Name:** `action`
- **Value Expression:** `login=='fail' ? 'blocked' : action`

Scenario C: Create a multivalued field called `myTags`. (i.e., array):

- **Name:** `myTags`
- **Value Expression:** `['failed', 'blocked']`

Scenario D: Add value `error` to the multivalued field `myTags`:

- **Name:** `myTags`
- **Value Expression:** `login=='error' ? [...myTags, 'error'] : myTags`

(The above expression is literal, and uses JavaScript spread syntax.)

Scenario E: Rename an `identification` field to the shorter `ID` – copying over the original field's value, and removing the old field:

- **Name**: `ID`
- **Value Expression**: `identification`
- **Remove Field**: `identification`

See Ingest-time Fields for more examples.

Usage Notes

Consider the following when working with Eval Functions.

Create Parent Objects First

Before you can use the Eval function on a new child object, you must create the parent object – then define the children using Eval Functions.

Example:

```
parent  =  (parent || { child1: child1Value })
parent.child2 = child2Value
parent.child3 = child3Value
<some other Evals, if you need them>
```

To Append:

```
parent =  Object.assign(parent, { child2: child2Value })
```

To Create a New Field:

```
parent2 =  Object.assign(parent, { child2: child2Value, child3: child3Value })
```

Execution Without Assignment

The Eval Function can execute expressions **without** assigning their value to the field of an event. You can do this by simply leaving the left-hand side input empty, and having the right-hand side do the assignment.

Example: Parse and Merge to Existing Field

`Object.assign(foo, JSON.parse(bar), JSON.parse(baz))` on the right-hand side (and left-hand side empty) will JSON-parse the strings in `bar` and `baz`, merge them, and assign their value to `foo`, an already existing field.

Example: Reference Event with `__e`

To parse JSON, enter `Object.assign(__e, JSON.parse(_raw))` on the right-hand side (and left-hand side empty). `__e` is a special variable that refers to the `(context)` event **within** a JS expression. In this case, content parsed from `_raw` is added at the top level of the event.

14.12. Event Breaker

This Function enables you to split large blobs or streams of events into discrete events within a Pipeline. This is useful for Sources like Azure Event Hubs, which do not natively support Event Breakers.

Even with Sources that do support Event Breakers (like Raw HTTP), it sometimes makes sense to use both a Source-configured Event Breaker on the incoming stream, and an Event Breaker Function within a Pipeline.

Limitations

The Event Breaker Function operates only on data in `_raw`. For other events, move the array to `_raw` and stringify it before applying this Function.

The largest event that this Function can break is about about 128 MB (`134217728` bytes). Events exceeding this maximum size will be split into separate events, but left unbroken. Cribl Edge will set these events' `__isBroken` internal field to `false`.

Unlike regular Event Breakers, Event Breaker Functions do not have names, only descriptions.

Usage

Use the following options to define this Event Breaker Function.

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Existing or New?: Whether to use an existing ruleset or create a new one. Defaults to `Use Existing`.

- When **Existing or New?** is set to `Use Existing`, the **Existing Ruleset?** drop-down appears and you choose a ruleset from there.
- When **Existing or New?** is set to `Create New`, the ruleset creation UI appears and you configure its settings.

Advanced Settings

Add to cribl_breaker: Whether to add the `cribl_breaker` field to output events. Defaults to `Yes`.

How this field behaves depends on whether you are **also** using a regular Event Breaker with your Source. That matters because a regular Event Breaker **always** adds the `cribl_breaker` field to events, which it does **before** the data reaches your Event Breaker Function.

When you are **not** using a regular Event Breaker, there is no `cribl_breaker` field yet when the data reaches your Event Breaker Function. At this point, `cribl_breaker` is added, and:

- When **Existing or New?** is set to `Use Existing`, `cribl_breaker`'s value is set to the name of the existing ruleset you chose.
- When **Existing or New?** is set to `Create New`, the `cribl_breaker`'s value is set to `"event_breaker_func"`. Since Event Breaker Functions do not have names, the string `event_breaker_func` serves as a kind of generic name that represents whatever new Event Breaker Function you created.

When you **are** also using a regular Event Breaker, the `cribl_breaker` field already exists when the data reaches your Event Breaker Function. At this point, the value of `cribl_breaker` is changed from a string to an array, with the first item in the array being the value originally set by the regular Event Breaker, and the second item determined as follows:

- When **Existing or New?** is set to `Use Existing`, the second item's value is set to the name of the existing ruleset you chose.
- When **Existing or New?** is set to `Create New`, the second item's value is set to `"event_breaker_func"`. Since Event Breaker Functions do not have names, the string `event_breaker_func` serves as a kind of generic name that represents whatever new Event Breaker Function you create.

Cribl Edge truncates timestamps to three-digit (milliseconds) resolution, omitting trailing zeros.

In Cribl Edge 3.4.2 and above, where an Event Breaker Function has set an event's `_time` to the current time – rather than extracting the value from the event itself – it will mark this by adding the internal field `__timestampExtracted: false` to the event.

Examples

Handling syslog data and nested JSON data are two primary use cases for Event Breaker Functions.

Event Breaker Functions for syslog Data

Event Breaker Functions can help you deal with syslog data that does not break correctly. Examples include multi-line syslog data, and, the kinds of non-standard syslog data emitted by Blue Coat proxy appliances, Layer7 API Gateways, and other appliances. See the Cribl video Scaling syslog for an in-depth discussion.

(In this context, "non-standard" means syslog data that only partly conforms to the standard Syslog Protocol defined in RFC 5424, or its predecessor, the BSD syslog Protocol defined in RFC 3164.)

Event Breaker Functions for Nested JSON Data

Kafka-based data from Confluent, Azure Event Hubs, Google Pub Sub, or Kafka itself, is nicely structured as events according to the Kafka protocol. But when these events contain JSON objects which you want to break into their constituent objects, you can use an Event Breaker Function to do that.

Troubleshooting

If you notice fragmented events, check whether Cribl Edge has added a `__timeoutFlush` internal field to them. This diagnostic field's presence indicates that the events were flushed because the Event Breaker buffer timed out while processing them. These timeouts can be due to large incoming events, backpressure, or other causes.

14.13. Flatten

The Flatten Function flattens fields out of a nested structure. It pulls up nested key-value pairs (fields) to a higher level in the object. You can specify:

- Individual fields to flatten.
- The depth at which to flatten fields.
- The delimiter to use when concatenating keys.
- An optional prefix to add to transformed field names.

The Flatten Function creates fully qualified names for promoted fields. If you simply need to promote fields without transforming their names, use the `eval` Function.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Fields: List of top-level fields to include for flattening. Defaults to an empty array, which means all fields. Limit to specific fields by typing in their names, separated by hard returns. Supports wildcards (`*`). Supports double-underscore (`__`) internal fields only if individually enumerated – not via wildcards.

Prefix: Prefix string for flattened field names. Defaults to empty.

Depth: Number representing the nested levels to consider for flattening. Minimum `1`. Defaults to `5`.

Delimiter: Delimiter to use for flattening. Defaults to `_` (underscore).

Example

Add the following test sample in **Preview** > **Paste a Sample**:

```
input
```

```
{
        "accounting": [{
                "firstName": "John",
                "lastName": "Doe",
                "age": 23
        }, {
                "firstName": "Mary",
                "lastName": "Smith",
                "age": 32
        }],
        "sales": [{
                "firstName": "Sally",
                "lastName": "Green",
                "age": 27
        }, {
                "firstName": "Jim",
                "lastName": "Galley",
                "age": 41
        }]
}
```

Under **Select Event Breaker**, choose **ndjson** (newline-delimited JSON), and click **Save as a Sample File**.

Here's sample output with all settings at default:

output

```
{
  "accounting_0_firstName": "John",
  "accounting_0_lastName": "Doe",
  "accounting_0_age": 23,
  "accounting_1_firstName": "Mary",
  "accounting_1_lastName": "Smith",
  "accounting_1_age": 32,
  "sales_0_firstName": "Sally",
  "sales_0_lastName": "Green",
  "sales_0_age": 27,
  "sales_1_firstName": "Jim",
  "sales_1_lastName": "Galley",
  "sales_1_age": 41,
}
```

Using the Flatten Function’s default settings, we successfully create top-level fields from the nested JSON structure, as expected.

14.14. GeoIP

The GeoIP Function enriches events with geographic fields, given an IP address. It works with MaxMind's GeoIP binary database.

Prerequisite

You need to host the `.mmdb` database file from MaxMind. The following steps cover this process at a high level. They link to our Managing Large Lookups topic, where you can find additional details.

1. Download and extract the `.mmdb` database file.
2. Determine where to place the database file. We recommend the `$CRIBL_HOME/state/` subdirectory, which is already listed in the default `.gitignore` file that ships with Cribl Edge.
   You always have the option to upload the file to Cribl Edge's Lookups Library. In a Cribl.Cloud deployment, this is currently the only option. For details, see Reducing Deploy Traffic.
3. Place the database file on your Edge Node(s). In a distributed deployment, we recommend having the file on the Leader and all Edge Nodes. Smaller deployments can get away with hosting the file only on the Leader Node.
4. Optionally, set up automatic updates of the database file.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

GeoIP file (.mmdb): Path to a MaxMind database, in binary format, with `.mmdb` extension.

If the database file is located within the lookup directory (`$CRIBL_HOME/data/lookups/`), the **GeoIP file** does not need to be an absolute path.

In distributed deployments, ensure that the MaxMind database file is in the same location on the Leader Node and all Edge Nodes. However, if you've uploaded `.mmdb` files via Cribl Edge's Lookups Library UI, just click this combo box to display and select them on a drop-down list. Your selection here will handle the path management automatically.

IP field: Field name in which to find an IP to look up. Can be nested. Defaults to `ip`.

Result field : Field name in which to store the GeoIP lookup results. Defaults to `geoip`.

Examples

Assume that you are receiving SMTP logs, and need to see geolocation information associated with IPs using the SMTP service.

Here's a sample of our data, from IPSwitch IMail Server logs:

```
03:19 03:22 SMTPD(00180250) [192.168.1.131] connect 74.136.132.88 port 2539 03:19 03:22 SMTPD(00180250) [74.136.132.88] EHLO msnbc.com 03:19 03:22 SMTPD(00180250) [74.136.132.88] MAIL FROM:<info-jjgcdshx@test.us> 03:19 03:22 SMTPD(00180250) [74.136.132.88] RCPT To:<user@domain.com>
```

In this example, we'll chain together three Functions. First, we'll use a Regex Extract Function to isolate the host's IP. Next, we'll use the GeoIP Function to look up the extracted IP against our geoIP database, placing the returned info into a new `__geoip` field. Finally we'll use an Eval Function to parse that field's city, state, country, ZIP, latitude, and longitude.

Function 1 – Regex Extract

Regex: `\[(?<ip>\S+)\]` Source field: `_raw` Result: `74.136.132.88`

Function 2 – GeoIP

Event's IP field: `ip` Result field: `__geoip`

Function 3 – Eval

Name	Value Expression
City	`__geoip.city.names.en`
Country	`__geoip.country.names.en`
Zip	`__geoip.postal.code`
Lat	`__geoip.location.latitude`
Long	`__geoip.location.longitude`

In the Eval Function's **Remove fields** setting, you could specify the `__geoip` field for removal, if desired. However, its `__` prefix makes it an internal field anyway.

For a hosted tutorial on applying the GeoIP Function, see Cribl's GeoIP and Threat Feed Enrichment Sandbox.

14.15. Grok

The Grok Function extracts structured fields from unstructured log data, using modular regex patterns.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Optional description of this Function's purpose in this Pipeline. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Pattern: Grok pattern to extract fields. Cick the Expand button at right to open a preview/valdiation modal. Syntax supported: `%{PATTERN_NAME:FIELD_NAME}`.

Click **Add pattern** to chain more patterns.

Source field: Field on which to perform Grok extractions. Defaults to `_raw`.

Management

You can add and edit Grok patterns via Cribl Edge's UI by selecting **Knowledge** > Grok Patterns.
Pattern files are located at: `$CRIBL_HOME/(default|local)/cribl/grok-patterns/`

Example

Example event:

```
{"_raw": "2020-09-16T04:20:42.45+01:00 DEBUG This is a sample debug log message"}
```

Pattern: `%{TIMESTAMP_ISO8601:event_time} %{LOGLEVEL:log_level} %{GREEDYDATA:log_message}` **Source Field**: `_raw`

Event after extraction:

```
{"_raw": "2020-09-16T04:20:42.45+01:00 DEBUG This is a sample debug log message",
  "_time": 1600226442.045,
  "event_time": "2020-09-16T04:20:42.45+01:00",
  "log_level": "DEBUG",
  "log_message": "This is a sample debug log message",
}
```

Note the new fields added to the event: `event_time`, `log_level`, and `log_message`.

References

- Syntax for a Grok pattern is `%{PATTERN_NAME:FIELD_NAME}`. E.g.: `%{IP:client} %{WORD:method}`.
- Useful link for creating and testing Grok patterns: http://grokconstructor.appspot.com
- Additional patterns are available here: https://github.com/logstash-plugins/logstash-patterns-core/tree/main/patterns

14.16. JSON Unroll

The JSON Unroll Function accepts a JSON object string `_raw` field, unrolls/explodes an **array** of **objects** therein into individual events, while also inheriting top level fields. See example(s). Cribl highly recommends not using this JSON Unroll function for certain types of data. Instead, perform the unrolling using an event breaker for those inputs which support configuring an event breaker. Specifying the event breaker type **JSON Array** and toggling the **JSON Extract Fields** option to **Yes** will accomplish the same unrolling but much more efficiently. This is recommended, for example, for CloudTrail and Office635 events, which are collected as JSON arrays.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Path: Path to array to unroll, e.g., `foo.0.bar`.

New name: The name that the exploded array element will receive in each new event. Leave empty to expand the array element with its original name.

Example(s)

Assume you have an incoming event that has a `_raw` field as a JSON object string like this:

Sample _raw field

```
{"date":"9/25/18 9:10:13.000 PM",
   "name":"Amrit",
   "age":42,
   "allCars": [
       { "name":"Ford", "models":[ "Fiesta", "Focus", "Mustang" ] },
       { "name":"GM", "models":[ "Trans AM", "Oldsmobile", "Cadillac" ] },
       { "name":"Fiat", "models":[ "500", "Panda" ] },
       { "name":"Blackberry", "models":[ "KEY2", "Bold Touch 9900" ] }
   ]
}
```


Settings:

Path: `allCars` **New Name**: `cars`

Output Events:

Resulting Events


```
Event 1:
{"_raw":"{"date":"9/25/18 9:10:13.000 PM","name":"Amrit","age":42,"cars":{"name":"I

Event 2:
{"_raw":"{"date":"9/25/18 9:10:13.000 PM","name":"Amrit","age":42,"cars":{"name":"(

Event 3:
{"_raw":"{"date":"9/25/18 9:10:13.000 PM","name":"Amrit","age":42,"cars":{"name":"I

Event 4:
{"_raw":"{"date":"9/25/18 9:10:13.000 PM","name":"Amrit","age":42,"cars":{"name":"I
```


Each element under the original **allCars** array is now placed in a **cars** field in its own event, inheriting original top level fields; **date**, **name** and **age**

14.17. Lookup

The Lookup Function enriches events with external fields, using lookup table files in CSV, compressed `.csv.gz`, or binary `.mmdb` format.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Lookup file path (.csv, .csv.gz): Path to the lookup file. Select an existing file that you've uploaded via Cribl Edge's UI at Knowledge > Lookups Library, or specify the path. You can reference environment variables via `$`, e.g.: `$CRIBL_HOME/file.csv`.

When you configure this field via a distributed deployment's Leader Node, Cribl Edge will swap `$CRIBL_HOME/groups/<groupname>/` for `$CRIBL_HOME` when validating whether the file exists. In this case, the default upload path changes from `$CRIBL_HOME/data/lookups` (single-instance deployments) to `$CRIBL_HOME/groups/<groupname>/data/lookups/` (distributed deployments).

Match mode: Defines the format of the lookup file, and indicates the matching logic that will be performed. Defaults to `Exact`.

Match type: For CIDR and Regex **Match mode**s, this attribute refines how to resolve multiple matches. `First match` will return the first matching entry. `Most specific` will scan all entries, finding the most specific match. `All` will return all matches in the output, as arrays. (Defaults to `First match`. Not displayed for Exact **Match mode**.)

Lookup fields (.csv): Field(s) that should be used to key into the lookup table.

- **Lookup field name in event**: Exact field name as it appears in events. Nested addressing supported.
- **Corresponding field name in lookup**: The field name as it appears in the lookup file. Defaults to the **Lookup field name in event** value. This input is optional.

Case-Sensitive / Multiple Matches

Lookups are case-sensitive by default. (See the **Ignore case** option below.)

> If the lookup file contains duplicate key names with different values, all **Match mode**s of this Function will use **only** the value in the key's **final** instance, ignoring all preceding instances.

Output field(s): Field(s) to add to events after matching the lookup table. Defaults to **all** if not specified.

- **Output field name from lookup**: Field name, as it appears in the lookup file.
- **Lookup field name in event**: Field name to add to event. Defaults to the lookup field name. This input is optional. Nested addressing is supported.

Advanced Settings

Reload period (sec): To periodically check the underlying file for mod-time changes, and reload the file if necessary, change the default `-1` value (disabled) to a positive integer representing the check interval in seconds.

> In distributed deployments, enabling a **Reload period** can generate conflicts with configuration updates, causing Pipelines to skip executing some Lookup Functions. Cribl recommends that you enable it only for lookup files not managed by Cribl Edge's UI, or lookup files that can be updated by an external process. (E.g., a threat list that you update via a cron job.)
>
> For lookup files that **are** managed by Cribl Edge's UI, a distributed Cribl Edge deployment will override this setting as necessary – skipping checks to prevent conflicts that could trigger skipped lookups. These restrictions do **not** apply to single-instance deployments.

Ignore case: Ignore case when performing **Match mode: Exact** lookups. Defaults to `No`.

Add to raw event: Whether to append the looked-up values to the `_raw` field, as key=value pairs. Defaults to `No`.

Examples

Example 1: Regex Lookups

Assign a `sourcetype` field to events if their `_raw` field matches a particular regex.

paloalto.csv

```
regex,sourcetype
"^[^,]+,[^,]+,[^,]+,THREAT",pan:threat
"^[^,]+,[^,]+,[^,]+,TRAFFIC",pan:traffic
"^[^,]+,[^,]+,[^,]+,SYSTEM",pan:system
```

Match mode: Regex

Match type: First match

When using the Lookup Function with Regex and `First match`, ensure that your lookup file contains no empty lines - not even at the bottom. Any empty row will cause the function to always return `true`.

Lookup field name in event: `_raw`

Corresponding field name in lookup: `regex`

Events before and after

```
### BEFORE:

{"_raw": "Sep 20 13:03:55 PA-VM 1,2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018/
{"_raw": "Sep 20 13:03:55 PA-VM 1,2018/09/20 13:03:58,FOOBAR,THREAT,end,2049,2018/(


### AFTER:

{"_raw": "Sep 20 13:03:55 PA-VM 1,2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018/
  "sourcetype": "pan:traffic"
  }
{"_raw": "Sep 20 13:03:55 PA-VM 1,2018/09/20 13:03:58,FOOBAR,THREAT,end,2049,2018/(
  "sourcetype": "pan:threat"
  }
```

Example 2: CIDR Lookups

Assign a `location` field to events if their `destination_ip` field matches a particular CIDR range.

paloaltoips.csv

```
range,location
10.0.0.0/24,San Francisco
10.0.0.0/16,California
10.0.0.0/8,US
```

Match mode: CIDR

Match type: See options below

Lookup field name in event: `destination_ip`

Corresponding field name in lookup: `range`

In **Match mode: CIDR** with **Match type: Most specific**, the lookup will implicitly search for matches from most specific to least specific. There is no need to pre-sort data.

Note that **Match mode: CIDR** with **Match type: First Match** is likely the most performant with large lookups. This can be used as an alternative to **Most specific**, if the file is sorted with the most specific/relevant entries first. This mode still performs a table scan, top to bottom.

Events before and after

```
### BEFORE:

{"_raw": "Sep 20 13:03:55 PA-VM 1, 2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018
  "destination_ip": "10.0.0.102"
  }

### AFTER with Match Type: First Match

{"_raw": "Sep 20 13:03:55 PA-VM 1, 2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018
  "destination_ip": "10.0.0.102",
  "location": "San Francisco"
  }

### AFTER with Match Type: Most Specific

{"_raw": "Sep 20 13:03:55 PA-VM 1, 2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018
  "destination_ip": "10.0.0.102",
  "location": "San Francisco"
  }

### AFTER with Match Type: All

{"_raw": "Sep 20 13:03:55 PA-VM 1, 2018/09/20 13:03:58,FOOBAR,TRAFFIC,end,2049,2018
  "destination_ip": "10.0.0.102",
  "location": [
    "San Francisco",
    "California",
    "US",
  ]}
```

More Examples and Scenarios

More examples:

- Ingest-time Lookups.
- Lookups and Regex Magic.
- Lookups as Filters for Masks.

See also:

- Managing Large Lookups to optimize file locations for large lookup files.
- Redis Function for faster lookups using a Redis integration.

14.18. Mask

The Mask Function masks, or replaces, patterns in events. This is especially useful for redacting PII (personally identifiable information) and other sensitive data.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Optionally, enter a simple description of this Function.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Masking Rules: This table defines pairs of patterns to match and replace. Defaults to empty. Click **Add Rule** to add each rule. Each row of the resulting table provides the following options:

- **Match Regex**: Pattern to replace. Supports capture groups. Use `/g` to replace all matches, e.g.: `/foo(bar)/g`
- **Replace Expression**: A JavaScript expression or literal to replace all matching content. Capture groups can be referenced with `g` and the group number – e.g., `g2` to reference the second capture group.
- **Enabled**: Toggle this off to disable matching individual rules, while retaining their configuration in the table. Useful for iterative development and debugging.

Apply to Fields: Fields on which to apply the masking rules. Defaults to `_raw`. Add more fields by typing in their names, separated by hard returns. Supports wildcards (`*`) and nested addressing. Supports double-underscore (`__`) internal fields only if individually enumerated – not via wildcards.

Negated terms are supported. When you negate field names, the fields list is order-sensitive. E.g., `!foobar` before `foo*` means "Apply to all fields that start with `foo`, except `foobar`." However, `!foo*` before `*` means "Apply to all fields, except for those that start with `foo`."

Advanced Settings

Evaluate fields: Optionally, specify fields to add to events in which one or more of the **Masking Rules** were matched. These fields can be useful in downstream processing and reporting. You specify the fields as key–value expression pairs, like those in the Eval Function.

- **Name**: Field name.

- **Value Expression**: JavaScript expression to compute the value (can be a constant).

Evaluating the Replace Expression

The **Replace expression** field accepts a full JS expression that evaluates to a value, so you're not necessarily limited to what's under `C.Mask`. For example, you can do conditional replacement: `g1%2==1 ? `fieldA="odd"` : `fieldA="even"``

The **Replace expression** can reference other event fields as `event.<fieldName>`. For example, `` `${g1}${event.source}` ``. Note that this is slightly different from other expression inputs, where event fields are referenced without `event.` Here, we require the `event.` prefix for the following reasons:

- We don't expect this to be a common case.
- Expanding the event in the replace context would have a high performance hit on the common path.
- There is a slight chance that there might be a `gN` field in the event.

The output of the Mask Function is a string even when the **Replace expression** produces a number.

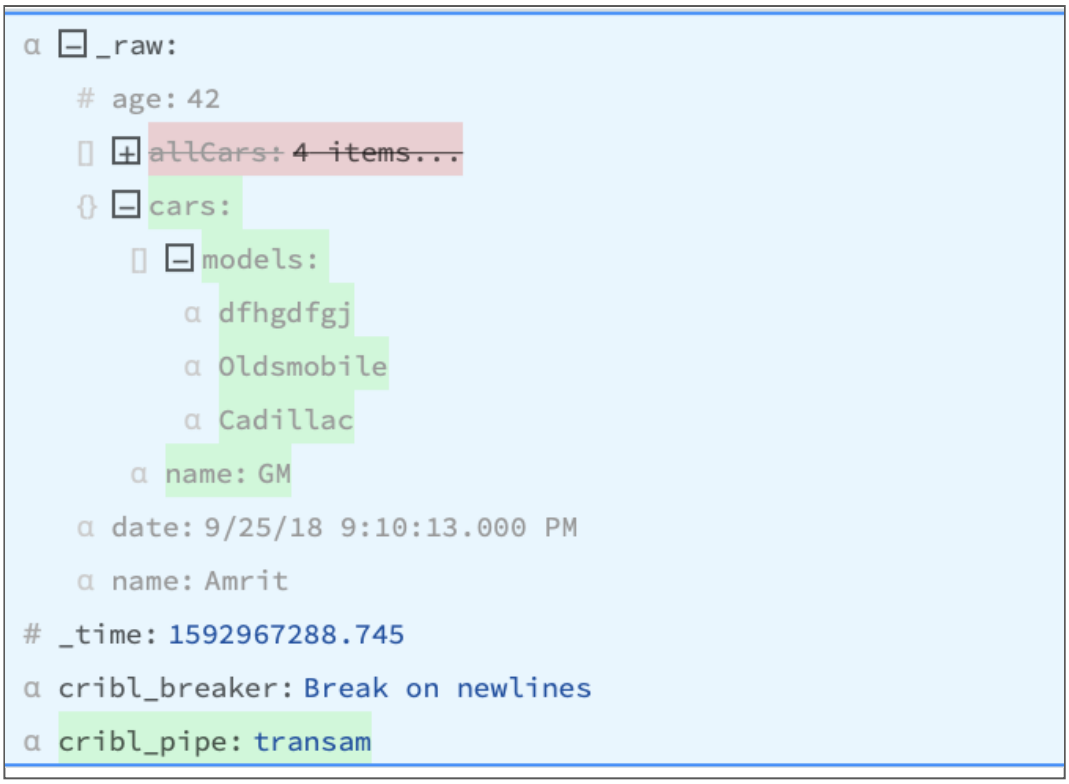
For example, suppose a **Replace expression** that generates a random integer is applied to a field named `bytes` and produces the number `8403`. Cribl Edge will output an event where the value of the `bytes` field is the **string** `8403`.

If you need to convert the string value back to a number, you could add a Numerify Function to your Pipeline.

Examples

Example 1: Transform a String

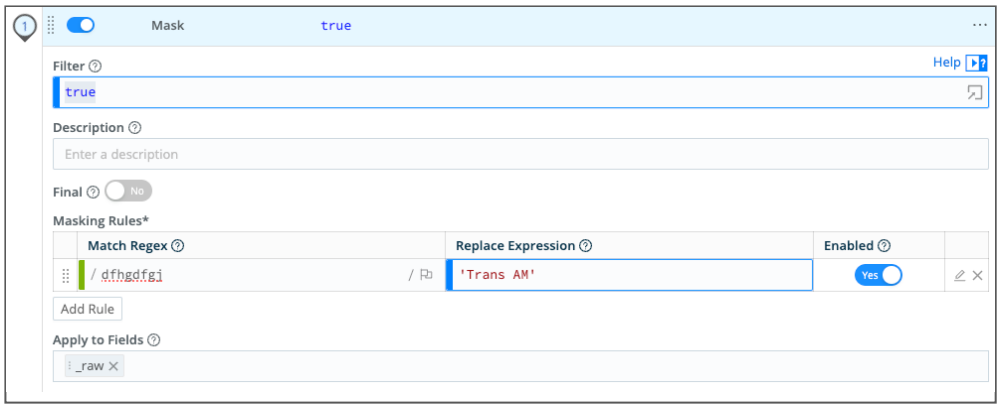
Here, we'll simply search for the string `dfhgdfgj`, and replace that value (if found) with `Trans AM`. This will help close America's muscle-car gap:



Event before masking

Configure the Mask Function > Masking Rules as follows:

Match Regex: `dfhgdfgj` Replace Expression: `Trans AM`



Mask Function configuration

Result: Vroom vroom!

Event after masking

Example 2: Mask Sensitive Data

Assume that you're ingesting data whose `_raw` fields contain unredacted Social Security numbers in the Key=Value pattern `social=#########`.

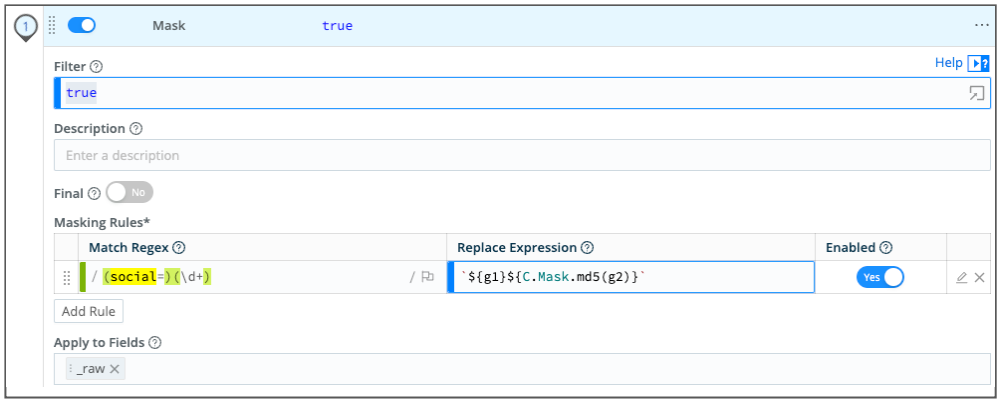
Event with unredacted SSNs

You can use a Mask Function to run an md5 hash of the `social` keys' numeric values, replacing the original values with the hashed values. Configure the Masking Rules as follows:

Match Regex: `(social=)(\d+)` Replace Expression: `` `${g1}${C.Mask.md5(g2)}` ``

In the first example everything in the Match regex field was replaced by the Replace Expression. However if that isn't desired then you can use capture groups in the Match Regex to define individual string components for manipulation or, alternatively, use string literals in the Replace expression for retaining any static text. Any content matching the Match Regex that is not inserted into the Replace expression will not be retained.

In this example, `social=` is assigned to capture group g1 for later reference. The value of `social=` will be hashed by referencing it as g2 in the md5 function. If we didn't make `social=` its own capture group (or specified `social=` as a literal in the Replace Expression) then we cannot reference it using g1 in the Replace expression, the value of `social=` would instead be assigned to g1, and the entire `social=#########` string would be replaced with a hash of the social security number, which probably isn't desired because no one would know the value being hashed without a field name preceding it.



Mask Function configuration

Result: The sensitive values are replaced by their md5 hashes.

Event with hashed SSNs

In scenarios where you need to send unmodified values to certain Destinations (such as archival stores), you can narrow the Mask Function's scope by setting the associated Route's **Output** field.

For further masking examples, see Masking and Obfuscation.

Example 3: Replace with an Event Field

In this example, we'll replace the IP address `127.0.0.1` in the `_raw` field with the IP address `192.168.123.25` from an existing field named `source_ip`. The following is a snippet of `_raw`:

```
ProcessId=0x0 IpAddress=127.0.0.1 IpPort=0
```

To match the IP address, we'll define three capture groups. Set **Match Regex** to:

```
/(IpAddress=)((?:\d{1,3}\.){3}\d{1,3})(\s)/
```

In the **Replace Expression**, we'll reference the `source_ip` field by prepending `event` to it, like this:

```
${g1}${event.source_ip}${g3}
```

Note that we've also referenced the first and third capture groups, using `g1` and `g3`. This changes `_raw` to:

```
ProcessId=0x0 IpAddress=192.168.123.25 IpPort=0
```

14.19. Numerify

The Numerify Function converts event fields that are numbers to type `number`.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Ignore fields: Specify fields to **not** numerify. Type in field names, separated by hard returns. Supports wildcards (`*`) and nested addressing. When empty (the default), Numerify applies to **all** fields. When populated, takes precedence over the **Include expression**.

Double Negatives

Ignore fields also supports negated terms. When you negate field names, the fields list is order-sensitive. E.g., `!foobar` before `foo*` means "Ignore all fields that start with `foo`, except `foobar`." However, `!foo*` before `*` means "Ignore all fields, except for those that start with `foo`."

Include expression: Optional JavaScript expression to specify fields to numerify. If empty (the default), the Function will attempt to numerify all fields – except those listed in **Ignore fields**, which takes precedence. Use the `name` and `value` global variables to access fields' names/values. (Example: `value != null`.) You can access other fields' values via `__e.<fieldName>`.

Format: Optionally, reformat or truncate the extracted numeric value. Select one of:

- **None**: Applies no reformatting (the default).
- **Floor**: Rounds the number down to the lower adjacent integer (truncates it).
- **Ceil**: Rounds the number up to the higher adjacent integer, removing decimal digits.
- **Round**: Rounds (truncates) the number to a specified number of digits. This option exposes an extra field:
    - **Digits**: Number of digits after the decimal point. Enter a value between `0`–`20`; defaults to `2`.

Examples

Scenario A:

Assume an event whose text contains a numeric value that must be extracted to perform some numeric analysis. The text looks like this:

```
version=11.5.0.0.1.1588476445
```

We can extract the numeric value by chaining together two Functions:

1. A Regex Extract Function. Set its **Regex** field to `/version=(?<ver>\d+)/`, to capture the first set of digits found in the event string.
2. Then use Numerify.

This captures the substring `11` and converts it to a numeric `11` value.

Scenario B:

Assume email transaction log events like the sample below. The final field is the message's size, in bytes. We want to extract this as a numeric value, for analysis in Cribl Edge or downstream services:

```
03:19 03:22 SMTPD (00180250) [209.221.59.70] C:\IMail\spool\D28de0018025017cd.SMD
3827
```

Again, we can accomplish this with two Functions:

1. A Regex Extract Function. To capture a substring of digits that follows six other substrings (all separated by white space), we set the **Regex** field to: `\S+\s+\S+\s+\S+\s+\S+\s+\S+\s+\S+\s+(?<bytes>\d+)`
2. Then use Numerify.

14.20. OTLP Metrics

The OTLP Metrics Function transforms dimensional metrics events into the OTLP (OpenTelemetry Protocol) format. This Function can be used to send metric events (events that contain the internal field `__criblMetrics`) to any OpenTelemetry Metrics-capable destinations.

For detailed insights into events that passed through the Function but were dropped, set the logging level for channel `func:otlp_metrics` to `debug`, and there will be a stats log message output once per minute with the relevant information.

This Function will delete the internal attribute `__criblMetrics`, making events only suitable for destinations that use the OTLP format.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. The default `true` setting passes all events through the Function.

Description: Simple description of this Function's purpose in this Pipeline. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Drop non-metric events: Whether or not to drop any non-metric events.

Basic Example

Filter: `__inputId=='prometheus_rw:prom_rw_in'`

This will grab all events from a Prometheus Remote Write Source and convert the metrics to OTLP format.

Ensure that the internal attribute `__criblMetrics` does not contain `null` or blank values in any of the child object arrays.

Sample event:

```
{
  "__criblEventType": "event",
  "__ctrlFields": [],
  "__final": false,
  "__cloneCount": 0,
  "_time": 1700236380.37,
  "_value": 123.456,
  "_metric": "cribl_logstream_health_outputs",
  "_metric_type": "gauge",
  "cribl_host": "P4MD6R-Pv8x",
  "cribl_wp": "w0",
  "event_host": "P4MD6R-Pv8x",
  "event_source": "cribl",
  "host": "P4MD6R-Pv8x",
  "output": "open_telemetry:otel_webhook_dest",
  "source": "cribl",
  "__criblMetrics": [
    {
      "nameExpr": ["_metric"],
      "values": ["_value"],
      "types": ["gauge"],
      "dims": [
        "cribl_host",
        "cribl_wp",
        "event_host",
        "event_source",
        "host",
        "output",
        "source"
      ]
    }
  ],
  "__offset": 8099,
  "__bytes": 228,
  "__srcIpPort": "127.0.0.1:50868",
  "__inputId": "prometheus_rw:prom_rw_in"
}
```


The same sample event after conversion by the OTLP Metrics Function:

```
{
  "_time": 1700236380.37,
  "instrumentation_library_metrics": [
    {
      "schema_url": "",
      "metrics": [
        {
          "__type": "gauge",
          "gauge": {
            "data_points": [
              {
                "attributes": [
                  {
                    key: "cribl_host",
                    value: {
                      string_value: "P4MD6R-Pv8x",
                    }
                  },
                  {
                    key: "cribl_wp",
                    value: {
                      string_value: "w0",
                    }
                  },
                  {
                    key: "event_host",
                    value: {
                      string_value: "P4MD6R-Pv8x",
                    }
                  },
                  {
                    key: "event_source",
                    value: {
                      string_value: "cribl",
                    }
                  },
                  {
                    key: "host",
                    value: {
                      string_value: "P4MD6R-Pv8x",
                    }
                  },
```

```
                {
                  key: "output",
                  value: {
                    string_value: "open_telemetry:otel_webhook_dest",
                  }
                },
                {
                  key: "source",
                  value: {
                    string_value: "cribl"
                  }
                },
              ],
              "start_time_unix_nano": 1700236380370000000,
              "time_unix_nano": 1700236380370000000,
              "as_double": 123.456
            }
          ]
        },
        "name": "cribl_logstream_health_outputs"
      }
    ]
  }
]
}
```

14.21. Parser

The Parser Function can be used to extract fields out of events, or to reserialize (rewrite) events with a subset of fields. Reserialization will maintain the format of the events.

For example: If an event contains comma-delimited fields, and `fieldA` and `fieldB` are filtered out, those fields' positions will be set to `null`, but not deleted completely.

Parser cannot remove fields that it did not create. A subsequent Eval Function can do so.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning evaluate all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Operation mode: **Extract** will create new fields. **Reserialize** will extract, filter fields, and then reserialize.

Type: Parser/Formatter type to use. Options:

- CSV – selecting this type requires you to escape certain characters.
- JSON
- K=V Pairs – selecting this type requires you to escape certain characters.
- Delimited Values – selecting this type expands the UI.
- Common Log Format (CLF)
- Extended Log File Format (ELFF)

The following two additional options appear in the drop-down, but can only be used when **Operation mode** is set to **Extract**; both expand the UI:

- Grok
- Regex Extract

Library: Select an option from the Parsers Library. Although specifying a Library will auto-generate an example list of fields, the list may still need to be modified to accommodate the desired fields from the events as well as the actual field order.

Source field: Field that contains text to be parsed. Not usually needed in Reserialize mode.

Destination field: Name of field in which to add extracted and serialized fields. If multiple new fields are created and this field is populated, then all new fields are created as elements of an array. The array's name will be set to this field's value. If you want all new fields to be independent, rather than in an array, then specify them using **List of fields**, below. (Extract and Reserialize modes only.)

Clean fields: This option appears for **Type: K=V Pairs**. Toggle to `Yes` to clean field names by replacing non-alphanumeric characters with `_`. This will also strip leading and trailing `"` symbols.

List of fields: Fields expected to be extracted, in order. If not specified, Parser will auto-generate fields.

Fields to keep: List of fields to keep. Supports wildcards (`*`). Takes precedence over **Fields to remove**. Supports nested addressing. See details below.

Fields to remove: List of fields to remove. Supports wildcards (`*`). Cannot remove fields matching **Fields to keep**. Supports nested addressing. See details below.

Fields filter expression: Expression to evaluate against each field's `{index, name, value}` context. Return truthy to keep fields, or falsy to remove fields. Index is zero-based. See details below.

Types That Expand the UI

Setting **Type** to **Delimited Values** displays the following extra options:

- **Delimiter**: Delimiter character to split value. Defaults to comma (`,`). You can also specify pipe (`|`) or tab characters.
- **Quote char**: Character used to quote literal values. Defaults to `"`.
- **Escape char**: Character used to escape delimiter or quote characters. Defaults to: `\`
- **Null value**: Field value representing the null value. These fields will be omitted. Defaults to: `-`

Setting **Type** to **Grok** displays the Grok Function UI; setting it to **Regex Extract** displays the Regex Extract UI. These Parser preserve all previewing and editing. Also, **Grok** enables you to bring Grok Patterns into Cribl Search.

Advanced Settings

Allowed key characters: Enter characters permitted in a key name, even though they are normally separator or control characters. Separate entries with a tab or hard return. This setting does not affect how the value is parsed.

Allowed value characters: Enter characters permitted in a value name, even though they are normally separator or control characters. Separate entries with a tab or hard return. This setting does not affect how the key is parsed.

Fields/Values Interaction

This Function's **Filter**, **Type**, **Fields to keep**, **Fields to remove**, and **Fields filter expression** settings interact as follows.

Order of Evaluation

The order is: **Fields to keep** > **Fields to remove** > **Fields filter expression**.

Precedence

If a field is in both **Fields to keep** and **Fields to remove**, **Fields to keep** takes precedence.

If a field is in both **Fields to remove** and **Fields filter expression**, **Fields to remove** takes precedence.

Negation

Both **Fields to remove** and **Fields to keep** support negated terms. When you use negated terms, your fields list is order-sensitive. E.g., `!foobar, foo*` means "All fields that start with `foo`, except `foobar`." However, `!foo*, *` means "All fields, except for those that start with `foo`."

Characters to Escape

Particular characters need special treatment when parsed when you select the Key=Value Pairs or CSV Parser/Formatter type.

Key-Value Pairs

When a Parser Function's **Type** is set to `Key=Value Pairs`: Within your data, field values that contain an `=` delimiter **must** be surrounded by quotes (`"..."`), or the Function will return unexpected results. As an example, here is a set of four fields and values, in which two values are quoted to parse predictably:

```
one=1,two="value=2",three=3,four="value=4"
```

These would resolve to the following parsed values:

Field Name	Value
one	1
two	"value=2"
three	3
four	"value=4"

CSV

When a Parser Function's **Type** is set to CSV: Within your data, column values that contain quotes or commas **must** quote those characters, or the Function will return unexpected results. To illustrate this, here is a properly formatted example of four CSV values:

```
"value with ""escaped"" quotes is 1",2,3,"you know, this is 4"
```

The first value includes quotation marks, which are each escaped with surrounding quotes, and the whole value is delimited by outer quotes. The fourth value includes a comma, so that value is delimited by outer quotes. These four columns would cleanly evaluate to:

---

value with "escaped" quotes is 1
2
3
you know, this is 4

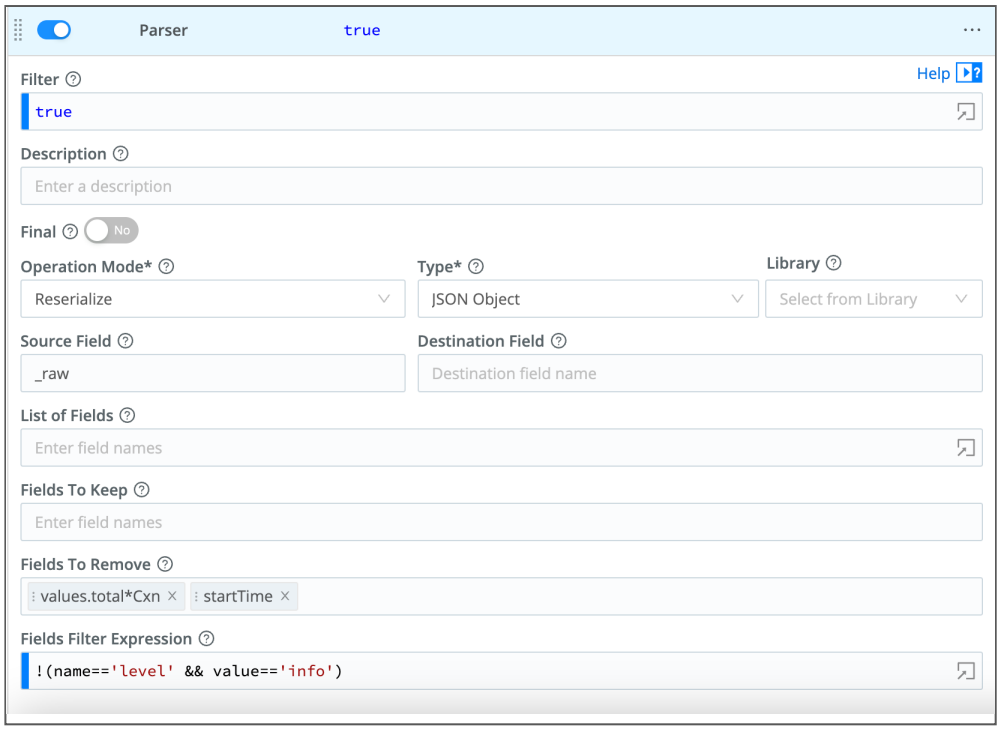
Examples

The following examples show creating and dropping fields, reserialization, and serializing to CSV versus JSON.

Example 1

Insert the following sample, using **Sample Data** > **Import Data**: `2019/06/24 05:10:55 PM Z a=000,b=001,c=002,d=003,e=004,f=005,g1=006,g2=007,g3=008`

Create the following test Parser Function (or import this Pipeline: https://raw.githubusercontent.com/weeb-cribl/cribl-samples/master/parser/functions/parser/parser_1.json).



Parser Function initial configuration

First, set the **Type** to `Key=Value Pairs`.

Scenario A

Keep fields `a, b, c`. Drop the rest.

Expected result: `a, b, c`

- Fields to Keep: `a, b, c`
- Fields to Remove: `*`
- Fields Filter Expression: <empty>

Result: The event will gain four new fields and values, as follows.

- a: `000`
- b: `001`
- c: `002`
- cribl_pipe: `parser2`

Scenario A result

You can check your stats by clicking the **Preview** pane's **Basic Statistics** (chart) button. In the resulting pop-up, the **Number of Fields** should have incremented ty four.

Now that you have the hang of it, try out the other simple scenarios below.

Scenario B

Keep fields `a`, `b`, those that start with `g`. Drop the rest.

Expected result: `a`, `b`, `g1`, `g2`, `g3`

- Fields to keep: `a`, `b`
- Fields to remove: [empty]
- Fields filter expression: `name.startsWith('g')`

Scenario C

Keep fields `a`, `b`, those that start with `g` but only if value is `007`. Drop the rest.

Expected result: `a`, `b`, `g2`

- Fields to keep: `a`, `b`
- Fields to remove: [empty]
- Fields filter expression: `name.startsWith('g') && value=='007'`

Scenario D

Keep fields `a`, `b`, `c`, those that start with `g`, unless it's `g1`. Drop the rest.

Expected result: `a, b, c, g2, g3`

- Fields to keep: `a, b, c`
- Fields to remove: `g1`
- Fields filter expression: `name.startsWith('g')`

Scenario E

Keep fields `a, b, c`, those that start with `g` but only if index is greater than `6`. Drop the rest.

Expected result: `a, b, c, g2, g3`

- Fields to keep: `a, b, c`
- Fields to remove: [empty]
- Fields filter expression: `name.startsWith('g') && index>6`

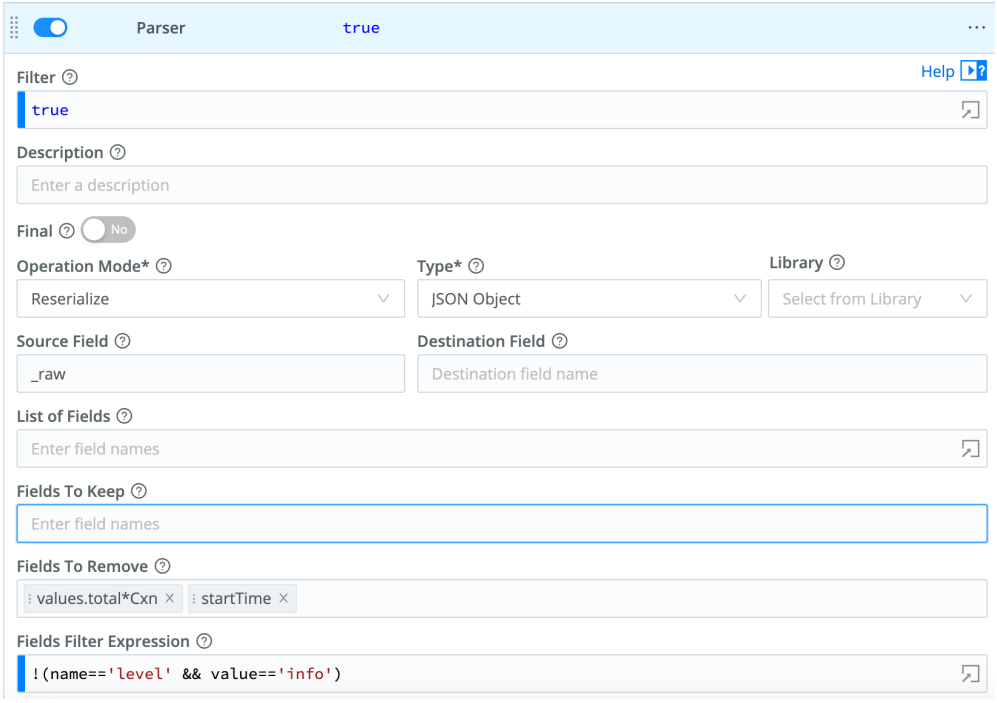
The `index` refers to the location of a field in the array of all fields extracted by **this** Parser. It is zero-based. In the case above, `g2` and `g3` have `index` values of `7` and `8`, respectively.

Example 2

Assume we have a JSON event that needs to be **reserialized**, given these requirements:

1. Remove the `level` field only if it's set to `info`.
2. Remove the `startTime` field, and all fields in the `values.total.` path that end in `Cxn`.

Parser Function configuration:



Parser Function configuration for Example 2

JSON event after being processed by the Function:

Example 2 event transformation

Example 3

Insert the following sample, using **Sample Data** > **Import Data**:

```
2019/06/24 15:25:36 PM Z a=000,b=001,c=002,d=003,e=004,f=005,g1=006,g2=007,g3=008,
```

For all scenarios below, first create a Parser Function to extract all fields, by setting the **Type** to `Key=Value Pairs`. Then add a second Parser Function with the configuration shown under **Parser 2**.

Scenario A

Serialize fields `a, b, c, d` in CSV format.

Expected result: `_raw` field will have this value `000,001,002,003`

Parser 2:

- Operation mode: `Serialize`
- Source field: [empty]
- Destination field: [empty]
- Type: `CSV`
- List of fields: `a, b, c, d` (needed for positional formats)

Scenario B

Serialize fields `a, b, c` in JSON format, under a field called `bar`.

Expected result: `bar` field will be set to: `{"a":"000","b":"001","c":"002","d":"003"}`

Parser 2:

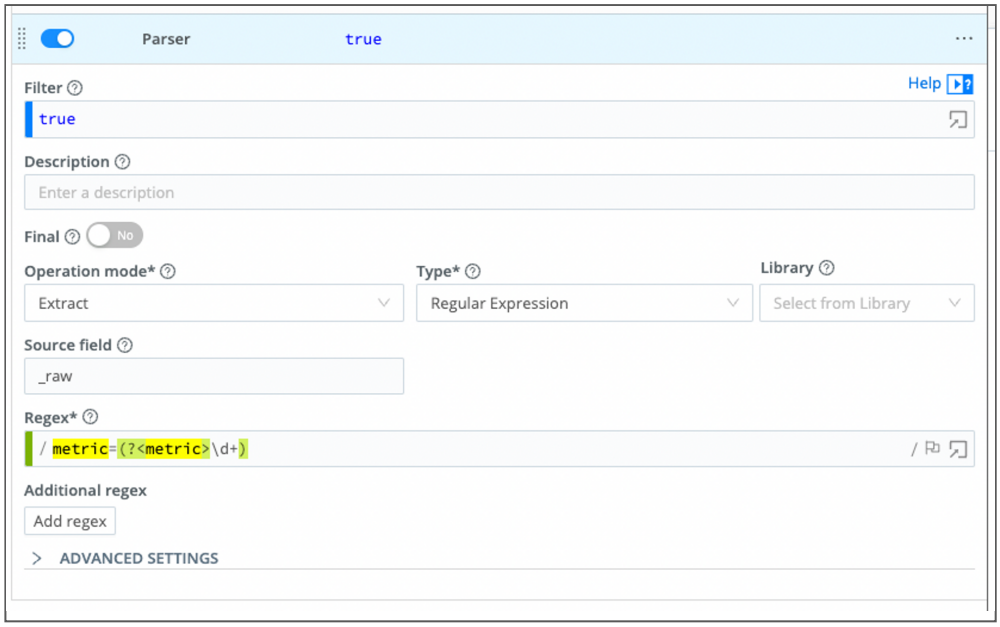
- Operation mode: `Serialize`
- Source field: [empty]
- Destination field: `bar`
- Type: `JSON`
- List of fields: [empty]
- Fields to keep: `a, b, c, d`

Example 4

Assume we have a simple event that looks like this:

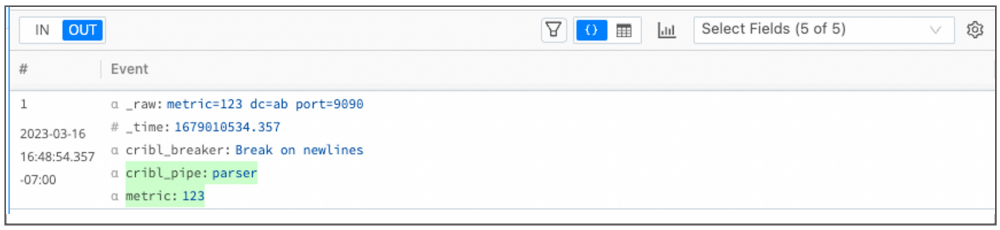
```
metric=123 dc=ab port=9090
```

If we want to extract only the `metric` field, we can configure a Parser of **Type** `Regular Expression`, where the **Regex** field is set to `metric=(?<metric>\\d+)`:



Parser Function configuration for Example 4

The output will look like this:



Parser Function regex output

Example 5

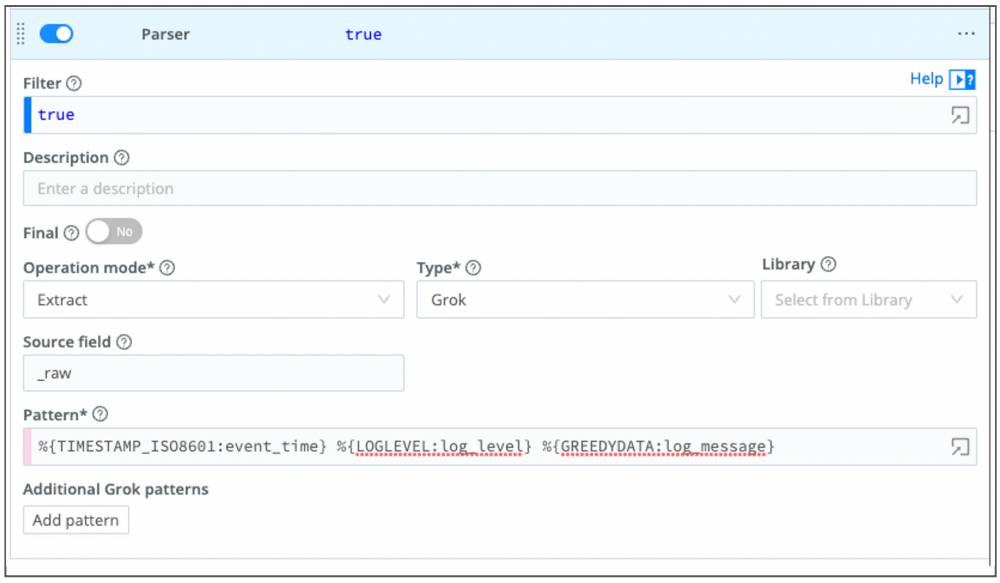
Assume we have an event that looks like this:

```
2020-09-16T04:20:42.45+01:00 DEBUG This is a sample debug log message
```

We can use a Grok pattern to parse the event so that the timestamp, log level, and the message are separated. A Parser of **Type** `Grok` with the following Pattern will work:

```
%{TIMESTAMP_ISO8601:event_time} %{LOGLEVEL:log_level} %{GREEDYDATA:log_message}
```

The Parser Function configuration looks like this:



Parser Function configuration for Example 5

The output will look like this:

Parser Function grok output

14.22. Publish Metrics

The Publish Metrics Function extracts, formats, and outputs metrics from events.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Overwrite: If set to `Yes`, overwrite previous metric specs. Otherwise, append. Defaults to `No`.

Metrics

Add Metrics: List of metrics to extract from the event and format. Destinations can pass the formatted metrics to a metrics aggregation platform. Click **Add Metrics** to add new rows containing the following options:

- **Event field name**: The name of the field (in the event) that contains the metric value. Should contain only letters, numbers, underscores (`_`), and `.` characters (to separate names in nested structures).
- **Metric name expression**: JavaScript expression to evaluate the metric field name. Defaults to the **Event field name** value.

The JavaScript expression will evaluate the metric field name only after the metrics are processed for transport to the Destination. While in the processing Pipeline, the metric name expression appears as a literal.

- **Metric type**: Select `Gauge` (the default), `Counter`, `Timer`, `Distribution`, `Summary`, or `Histogram`. General definitions that can vary across senders:
  - `Gauge`: A numeric value that can increase or decrease over time – like a temperature or pressure gauge.
  - `Counter`: A cumulative numeric value – it can only increase over time.
  - `Timer`: Generally measures how long a given event type takes (duration), and how often it occurs (frequency).

- `Distribution`: The statistical distribution of a set of values over a time interval. (This type generally provides raw data, not an aggregation.)
- `Summary`: Tracks the count, sum, and average of the values. Optionally, also tracks quantiles across the values.
- `Histogram`: Tracks the count, sum, and average of the values. Also groups the observations in their corresponding intervals, or buckets.

Remove Metrics: Optionally, enter a List of field names to look for when removing metrics. Where a metric's field name matches an element in this list, Cribl Edge will remove that metric from the event.

Dimensions

Add Dimensions: Optional list of dimensions to include in events. Supports wildcards. If you don't specify metrics, values will be appended to every metric found in the event. When you add a new metric, dimensions will be present only in those new metrics. Defaults to `!_* *`.

Remove Dimensions: Optional list of dimensions to associate with every extracted metric value. Leave blank if this Function processes output from the Aggregations Function as dimensions will be automatically discovered.

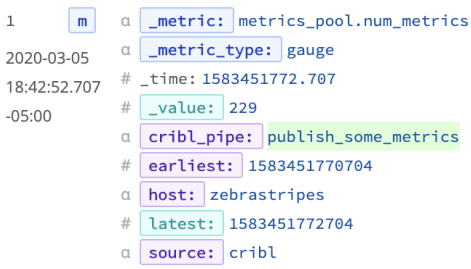
Overwrite: If set to `Yes`, overwrite previous metric specs. Otherwise, append. Defaults to `No`.

The **Add Dimensions** and **Remove Dimensions** fields support wildcards and negated terms. When you use negated terms, the list is order-sensitive. E.g., `!foobar` before `foo*` means "All fields that start with `foo`, except `foobar`." However, `!foo*` before `*` means "All fields, except for those that start with `foo`."

Cribl Edge can send out multi metrics, but this must be configured on compatible Destinations.

Fields Color Coding

On the right Preview pane's **OUT** tab, the Publish Metrics Function adds the following color codes to field labels:



Dimension: purple | **Value**: cyan (light blue) | **Info**: dark blue

These are in addition to the color codes applied to field values, which are listed here.

Examples

Scenario A:

Assume we're working with AWS VPC Flowlog events that have the following structure:

```
version account_id interface_id srcaddr dstaddr srcport dstport protocol packets
bytes start end action log_status
```

For example:

```
2 99999XXXXX eni-02f03c2880e4aaa3 10.0.1.70 10.0.1.11 9999 63030 6 6556 262256
1554562460 1554562475 ACCEPT OK
```

... and we want to use values of `packets` and `bytes` as metrics across these dimensions: `action`, `interface_id`, and `dstaddr`.

To reference the `packets` and `bytes` fields by name, as `'packets'` and `'bytes'`, our Pipeline will need a Parser Function before the Publish Metrics Function.

Parser Function

Filter: Set as needed Operation mode: Extract Type: Extended Log File Format (automatically set when specifying a library) Library: AWS VPC Flow Logs Source: `_raw` (No need to specify any other fields.)

Publish Metrics Function

Below, the `metric_name` prefix was arbitrarily chosen. Because there is no JavaScript expression to evaluate – i.e., this is literal text – the strings specified for the **Metric name expression** will be identical to those in the final metrics data sent to the Destination. See Raw Output below.

Metrics

Event Field Name	Metric Name Expression	Metric Type
`bytes`	`` `metric_name.bytes` ``	**Gauge**

Event Field Name	Metric Name Expression	Metric Type
packets	`metric_name.packets`	Gauge

Dimensions

```
action interface_id dstaddr
```

All specified dimension names must align with those from the original event. When you preview the Function's output, the metrics and dimensions will all have special highlighting to separate them from other fields. Additional highlighting is used to differentiate the metrics from the dimensions. (If one or more metrics/dimensions are not highlighted as expected, check the Function's configuration.)

Raw Output

```
metric_name.bytes:262256|g#action:REJECT,interface_id:eni-
02f03c2880e4aaa3,dstaddr:10.0.1.11

metric_name.packets:6556|g#action:REJECT,interface_id:eni-
02f03c2880e4aaa3,dstaddr:10.0.1.11
```

Compatible Destinations

All text after the `#` symbol represents the dimensions as key-value pairs. In order for dimension data to be included in metrics, the Destination type cannot be standard **StatsD**. However, **StatsD Extended**, **Splunk**, and **Graphite** do support dimensions.

Formatted Output

```
{
  "action": "REJECT",
  "interface_id": "eni-02f03c2880e4aaa3",
  "dstaddr": "10.0.1.11",
  "metric_name.bytes": 262256,
  "metric_name.packets": 6556,
}
```

Scenario B:

Assume that we want to extract some metrics from specific fields in PANOS logs, whose events have the following structure:

```
future_use_0,receive_time, serial_number, type, threat_content_type, future_use_1,
generated_time, source_ip, destination_ip, nat_source_ip, nat_destination_ip,
rule_name, source_user, destination_user, application, virtual_system, source_zone,
destination_zone, inbound_interface, outbound_interface, log_action, future_use_2,
session_id, repeat_count, source_port, destination_port, nat_source_port,
nat_destination_port, flags, protocol, action, bytes, bytes_sent, bytes_received,
packets, start_time, elapsed_time, category, future_use_3, sequence_number,
action_flags, source_location, destination_location, future_use_4, packets_sent,
packets_received, session_end_reason, device_group_hierarchy_level_1,
device_group_hierarchy_level_2, device_group_hierarchy_level_3,
device_group_hierarchy_level_4, virtual_system_name, device_name, action_source,
source_vm_uuid, destination_vm_uuid, tunnel_id_imsi, monitor_tag_imei,
parent_session_id, parent_start_time, tunnel_type, sctp_association_id, sctp_chunks,
sctp_chunks_sent, sctp_chunks_received
```

For example:

```
Jan 10 10:19:15 DMZ-internal.nsa.gov 1,2019/01/10
10:19:15,001234567890002,TRAFFIC,drop,2304,2019/01/10
10:19:15,209.118.103.150,160.177.222.249,0.0.0.0,0.0.0.0,InternalServer,,,not-
applicable,vsys1,inside,z1-FW-Transit,ethernet1/2,,All traffic,2019/01/10
10:19:15,0,1,63712,443,0,0,0x0,udp,deny,60,60,0,1,2019/01/10
10:19:15,0,any,0,0123456789,0x0,Netherlands,10.0.0.0-10.255.255.255,0,1,0,policy-
deny,0,0,0,0,,DMZ-internal,from-policy,,,0,,0,,N/A,0,0,0,0,1202585d-b4d5-5b4c-aaa2-
d80d77ba456e,0
```

Our goal is to use the four values of `bytes_sent`, `bytes_received`, `packets_sent`, and `packets_received` as metrics across these dimensions: `destination_ip`, `inbound_interface`, `outbound_interface`, and `destination_port`.

Here again, our Pipeline will need a Parser Function before the Publish Metrics Function.

Parser Function

Filter: Set as needed Operation mode: Extract Type: Extended Log File Format (automatically set when specifying a Library) Library: Palo Alto Traffic Source: `_raw` (No need to specify any other fields.)

Publish Metrics Function

Set up the Publish Metrics Function as follows.

Metrics

Event Field Name	Metric Name Expression	Metric Type
`bytes_sent`	`` `metric.${host}.bytes_sent` ``	Counter
`bytes_received`	`` `metric.${host}.bytes_rcvd` ``	Counter
`packets_sent`	`` `metric.${host}.pkts_sent` ``	Counter
`packets_received`	`` `metric.${host}.pkts_rcvd` ``	Counter

Added Dimensions

`destination_ip`, `inbound_interface`, `outbound_interface`, `destination_port`

Raw Output

```
metric.10.10.12.192.bytes_sent:60|c|#destination_ip:160.177.222.249,inbound_interface:
metric.10.10.12.192.bytes_rcvd:0|c|#destination_ip:160.177.222.249,inbound_interface:e
metric.10.10.12.192.pkts_sent:1|c|#destination_ip:160.177.222.249,inbound_interface:et
metric.10.10.12.192.pkts_rcvd:0|c|#destination_ip:160.177.222.249,inbound_interface:et
```

Here again, all text after the `#` symbol represents the dimensions as key-value pairs. (See the Compatible Destinations note above.) Unlike the first example, this example uses JavaScript expressions, which you can see evaluated in the raw output where the `${host}` has been converted to `10.10.12.192`.

14.23. Redis

The Redis Function interacts with Redis stores, setting and getting key-hash and key-value combinations. Redis' in-memory caching of these key pairs enables large lookup tables that would be cumbersome with a CSV or binary lookup file.

You can use Cribl Edge Collectors (for example, a REST Collector) to retrieve reference data from desired endpoints, and then use this Function to store the data on Redis and retrieve it to enrich your data. By default, Cribl Edge does not cache the data returned from this Redis Function, but you can enable client-side caching.

Your Redis Function can interact with three kinds of Redis instances: Standalone (the default); Redis Cluster, which scales horizontally; or Redis Sentinel, a distributed system that provides high availability.

Managing Redis Connections

When using Redis Functions, you must consider how to avoid creating large numbers of TCP connections to the Redis server (which we'll call "Redis connections"). To manage Redis connections, you start by calculating how many Redis connections your Redis Functions, other functions, and Pipelines will create using default settings. Then, to keep the overall number of connections reasonable, you configure Cribl Edge to **reuse** Redis connections.

What causes the number of Redis connections to grow large? By default, Cribl Edge creates a separate TCP connection to Redis for every Redis Function and every **reference to** a Redis Function. This means that certain usage patterns will increase the number of Redis Connections, including: sending the output of a Chain Function to a regular Pipeline that includes a Redis Function; using a Redis Function in a Source's pre-processing Pipeline; and, using a Redis Function in a Destination's post-processing Pipeline.

How can you control the number of Redis connections created? You use the **Reuse Redis connections** control. Then, when a new Redis connection is needed, **and** its connection properties are identical to those of an existing Redis connection, Cribl Edge will try to reuse the existing connection instead of creating a new one. The example below illustrates how powerfully this can affect the demands Cribl Edge places on a Redis server.

What are the connection properties that must be identical for reuse to be possible? For the different kinds of Redis instances, the relevant connection properties are, respectively:

- Standalone: username, password, and URL.
- Cluster: username, password, and all root node hosts and ports.
- Redis Sentinel: username, password, master name, and all root node hosts and ports.

In managing Redis connections, the factors you must account for include: what kind of Redis instance you're connecting to (Standalone, Cluster, or Redis Sentinel); the size of the Redis server; the number of connections the Redis server is configured to allow; the number of other clients using the same Redis server; and so on.

Where to configure **Reuse Redis connections** in the UI depends on how you have deployed Cribl Edge:

- Distributed deployment: Navigate to **Limits** at the Worker Group level.
- Single-instance deployment: Navigate to **Limits** at the global (**General Settings**) level.

Example: Managing Redis Connections

Referencing a Redis Function from a Pipeline and/or another Function can dramatically multiply the number of Redis connections created, unless we mitigate this effect by applying the **Reuse Redis connections** control. Let's see how this plays out in a detailed example. Suppose we do the following:

1. Create a Pipeline (we'll call it "Pipeline Zero") that contains 10 Redis Functions that connect to the same Redis server, with identical connection properties. Cribl Edge will create 10 TCP connections to the Redis server ("Redis connections").
2. Reference Pipeline Zero in a Chain Function that's in another Pipeline. This creates 10 more connections, for a total of 20 Redis connections.
3. Reference Pipeline Zero in an additional, separate Chain Function that's in yet another Pipeline, creating 10 more connections, for a total of 30 Redis connections.
4. Use Pipeline Zero as a pre-processing Pipeline for a Source, creating 10 more connections, for a total of 40 Redis connections.
5. Use Pipeline Zero as a post-processing Pipeline for a Destination, creating 10 more connections. Now we have a grand total of 50 Redis connections.

Now consider the effect of these Pipelines within Edge Nodes and their Worker Processes. For a typical deployment scenario with 10 Edge Nodes, each of which have 10 Worker Processes, the calculation looks like this:

- 50 connections times 10 Worker Processes = 500 connections, times 10 Edge Nodes = 5000 connections.

A single Redis server can be hard-pressed to handle so many connections. Apart from that, the number of connections is unnecessarily high for the expected throughput. Let's use **Reuse Redis connections** to reduce the number of Redis connections.

Enable **Reuse Redis connections** and set **Max connections** to `4`. Now revise the calculation:

- 4 connections times 10 Worker Processes = 40 connections, times 10 Edge Nodes = 400 connections.

This reduces the number of connections by better than a factor of 10, making our deployment much less likely to overwhelm the Redis server.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Result field: Name of the field in which to store the returned value. (Leave empty to discard the returned value.)

Command: Redis command to perform. Required. (A complete list of Redis commands is at: https://redis.io/commands.)

Key: A JavaScript expression to compute the value of the key to operate on. Can also be a constant such as `username`. This is a required field. Click the icon at right to open a validation modal.

Args: A JavaScript expression to compute arguments to the operation. Click the icon at right to open a validation modal. When the expression you enter is a plain value or an array, it resembles the options and arguments of a Redis command in the Redis CLI.

For example:

- Suppose you want to `SET` a key named `_mascot`. And, of the options that the `SET` command takes, you want to use two: `GET` and `EX`. The first, `GET`, takes no arguments; the second, `EX`, takes an integer in seconds.
- You want to set `_mascot` to the value `goat`, and `EX` to 5 seconds. Here is the array you'd enter in the Args column:
  `['goat', 'GET','EX',5 ]`

Deployment Type

From the drop-down, select `Standalone`, `Cluster`, or `Sentinel`. Some of these options display additional controls below. Hostnames must be JavaScript expressions (which can evaluate to a constant value), enclosed in quotes or backticks. They can be evaluated only at init time.

Standalone

This is the default deployment type. Enter a **Redis URL** to connect to. The format is:
`[redis[s]:]//[[user][:password@]][host][:port][/db-number][?db=db-number[&password=bar[&option=value]]]`

Example:
`redis://user:secret@localhost:6379/0?foo=bar&qux=baz`

Example with no `user` specified:
`redis://secret@localhost:6379/0?foo=bar&qux=baz`

Standalone with TLS

To enable TLS, specify a Redis URL that starts with `rediss://` rather than `redis://`. This URL reveals the TLS section of the config UI.

Example:
`rediss://user:secret@localhost:6379/0?foo=bar&qux=baz`

Example with no `user` specified:
`rediss://secret@localhost:6379/0?foo=bar&qux=baz`

Cluster

Specify **Root Nodes** that the cluster will connect to. A single node is sufficient, but Cribl recommends listing two or more. Each of the **Root Nodes** consists of a **Hostname** and a **Port**.

In multi-key commands, ensure that all keys in the command are in the same hash slot. Otherwise, the command will fail. As a workaround, you can split the multi-key command into separate commands, where each command operates only on keys from a single hash slot.

Sentinel

Specify a **Master Group Name**, plus every **Sentinel** in your Redis deployment. Each **Sentinel** consists of a **Hostname** and a **Port**.

Authentication Method

Use the **Authentication method** buttons to select one of the following options, some of which display additional controls below.

- **None**: Select this option where authentication either is not required, or is provided in the URL.

- **Basic**: This displays **Username** and **Password** fields for you to enter your Redis credentials.
- **User Secret**: This option exposes a drop-down in which you can select a stored text secret that references a Redis username and password, as described above. A **Create** link is available to store a new, reusable secret.
- **Admin Secret**: This option exposes a drop-down in which you can select a stored text secret that references a Redis admin password. A **Create** link is available to store a new, reusable secret.

TLS

Validate server certs: Reject certificates that are **not** authorized by a CA in the **CA certificate path**, nor by another trusted CA (such as the system's CA). Defaults to `Yes`.

Server name (SNI): Server name for the SNI (Server Name Indication) TLS extension. This must be a host name, not an IP address.

Certificate name: The name of the predefined certificate.

CA certificate path: Path on client containing CA certificates (in PEM format) to use to verify the server's cert. Path can reference `$ENV_VARS`.

Private key path (mutual auth): Path on client containing the private key (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Certificate path (mutual auth): Path on client containing certificates (in PEM format) to use. Path can reference `$ENV_VARS`. **Use only if mutual auth is required**.

Passphrase: Passphrase to use to decrypt private key.

Minimum TLS version: Optionally, select the minimum TLS version to use when connecting. The minimum TLS version for Redis is version 6.

Maximum TLS version: Optionally, select the maximum TLS version to use when connecting.

Minimum TLS Requirement

The minimum Redis version required for TLS support is version 6.

Advanced Settings

Max blocking time: Maximum amount of time (in seconds) before assuming that Redis is down and passing events through. Defaults to `60` seconds. Use `0` to disable timeouts.

Client-side cache: Toggle to `Yes` to enable. Then navigate to **Settings** > **General Settings** > **Limits** > **Redis Cache** to configure the cache, as described in the next section.

Overall Redis Cache Settings

Together, these settings determine how the Cribl Edge client-side cache behaves. We'll start with definitions:

Key TTL in seconds: The maximum time-to-live of a key in the cache, in seconds. `0` means no limit. Defaults to 10 minutes.

Max # of keys: The maximum number of keys the cache can contain. `0` (the default) means no limit.

Max cache size (bytes): The maximum number of bytes the cache can contain. `0` (the default) means no limit.

Service period (seconds): The interval at which the cache checks whether any keys need to be evicted. Defaults to 30 seconds.

Server assisted: This feature requires Redis version 6 or newer. Toggle to `Yes` to enable server-assisted caching and display the **Client tracking mechanism** drop-down. See the Tradeoffs section for an explanation of the available options.

Now here's the context:

For each Redis instance your Redis Functions interact with, Cribl Edge will operate one client-side cache. That cache will be shared between all the Redis Functions instances configured to interact with it. This is why the settings are in a central location, and why you cannot configure the cache at the individual function level.

Once per **Service period**, the cache checks whether any keys need to be evicted. The cache will evict:

- The least-recently-accessed key, if the cache has reached either (1) its **Max cache size (bytes)** or (2) its **Max # of keys**.
- Any key that exceeds its **Key TTL in seconds**.
- Any key that is stale according to a notification to Cribl Edge from your Redis instance. Redis sends these notifications (a.k.a. invalidations, or invalidation messages) only when you have enabled **Server assisted** caching.

The overall benefit that client-side caching provides is having fewer stale keys in your cache for less time. The greater the proportion of `GET`-type, as compared to `SET`-type (write) commands your Cribl Edge-Redis interactions have, the greater that benefit will be. When your Cribl Edge Pipelines either have no `SET`-type operations, or use filters to keep those operations to a minimum, you should consider client-side caching. In sum, you must decide whether more accurate data in your cache is worth the cost of memory and processing.

Tradeoffs Inherent in Server-Assisted Caching

Consider the common scenario where other applications besides Cribl Edge are writing to your Redis instance. Say one of those applications modifies a key's value. Now the value that Cribl Edge client-side cache has for that key is stale. Cribl Edge does not "know" about this, and retains the stale value until its TTL is up. This degrades the overall accuracy of data from the client-side cache.

Server-assisted caching is a remedy for this problem. With server-assisted caching enabled, your Redis instance will notify your client-side cache when Redis determines that a key's value has changed. Then the Cribl Edge client-side cache will immediately evict the stale key. Now the client-side cache is more accurate. Redis also has documentation about server-assisted client-side caching.

There are two mutually-exclusive ways you can have Redis notify Cribl Edge. These are the two modes (delivery mechanisms) available in the **Client tracking mechanism** drop-down:

1. In **Default** mode, the Redis server remembers which keys the Cribl Edge client has requested. Redis only notifies Cribl Edge when the value of one of **those** keys has changed. This is most efficient for Cribl Edge but needs more memory and some extra processing time from the Redis server.
2. In **Broadcast** mode, the Redis server need not remember which keys came from where, because it simply notifies Cribl Edge when the value of **any** key has changed. This is most efficient for Redis, but now the the Cribl Edge client needs to inspect every key that Redis says has changed. This consumes some extra processing time on the client, and generates extra network activity to accommodate notifications about keys that Cribl Edge never sees.

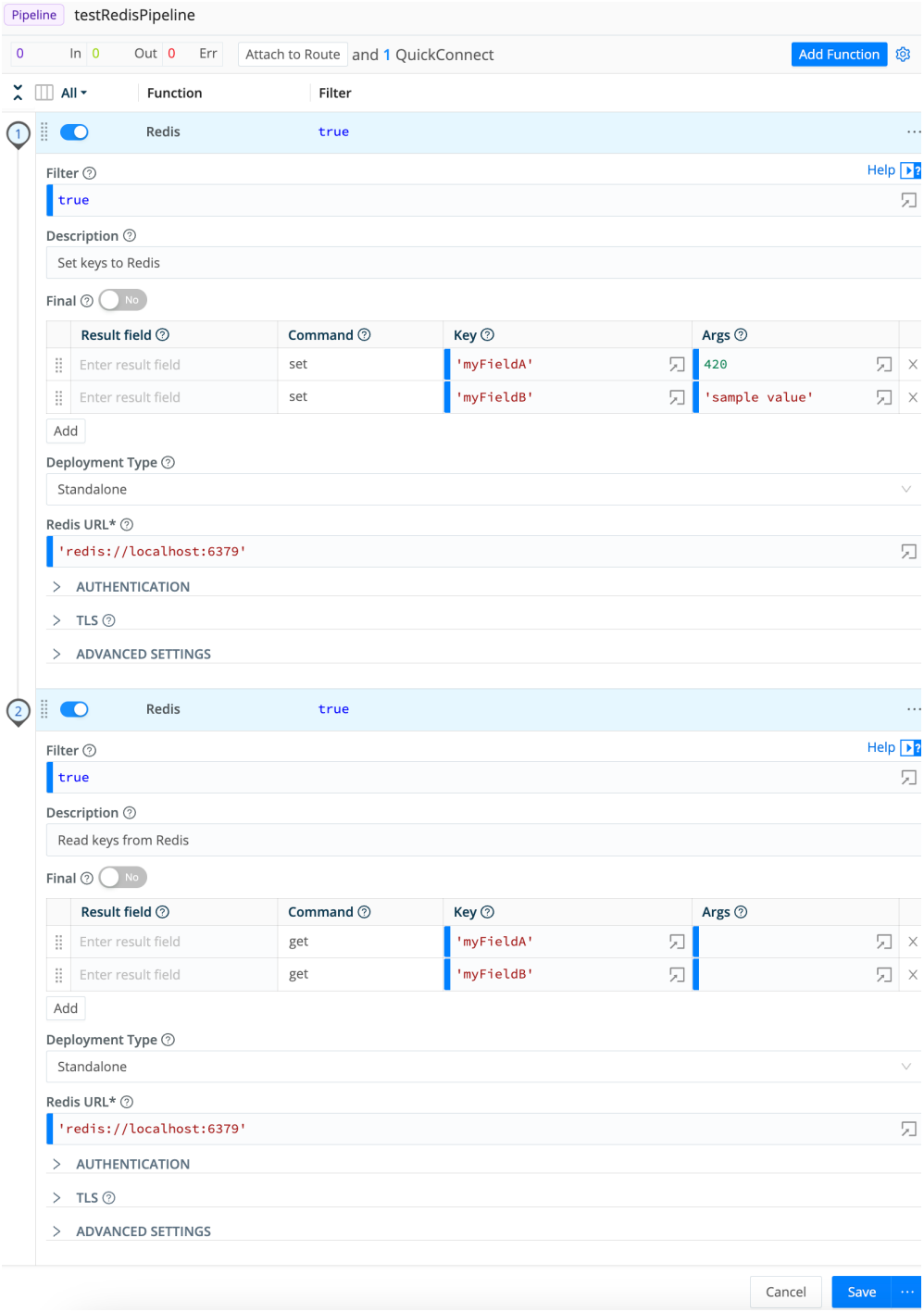
You should select whichever of these mechanisms best suits your environment. Does it make more sense to impose the heavier burden on the Redis server (as in **Default** mode) or on Cribl Edge (as in **Broadcast** mode)?

In both scenarios, for a given key, any write operation from any application (including Cribl Edge) that's interacting with Redis, causes Cribl to evict the key from the client-side cache when Redis notifies Cribl.

Examples

Scenario A: Set and Get

This Pipeline demonstrates the use of a pair of Redis Functions. The first Function sets two key-value pairs in Redis. The second Function gets their values, by key, into two corresponding new **Result field**s.



Redis set and get Functions

Redis Function #1

Description: `Set keys to Redis`

Command: `set` **Key**: `'myFieldA'` **Args**: `420`

Command: `set` **Key**: `'myFieldB'` **Args**: `'sample value'`

Redis Function #2

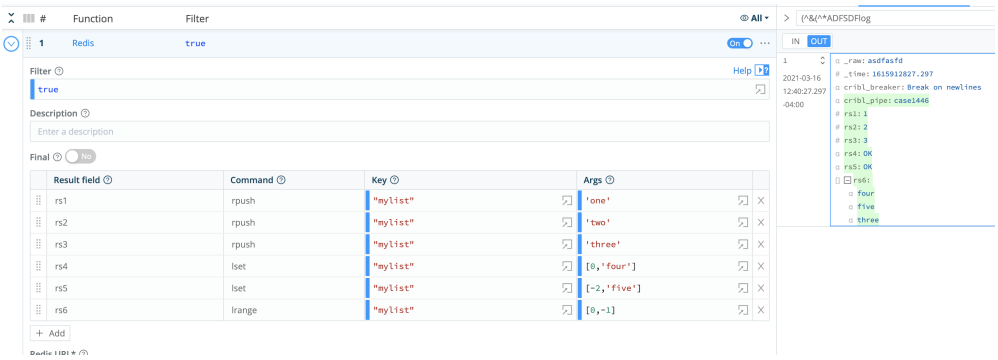
Description: `Read keys from Redis`

Result field: `myField_AA` **Command**: `get` **Key**: `'myFieldA'`

Result field: `myField_BB` **Command**: `get` **Key**: `'myFieldB'`

Scenario B: Multiple-Argument Arrays

This example demonstrates how to configure a Redis Function that supplies an array of multiple arguments to Redis commands (in this example, `lset` and `lrange`).



Redis Function, arrays of multiple arguments, and sample output

Redis Function

Description: `Push arrays of multiple arguments to Redis`

Result field: `rs1` **Command**: `rpush` **Key**: `"mylist"` **Args**: `'one'`

Result field: `rs2` **Command**: `rpush` **Key**: `"mylist"` **Args**: `'two'`

Result field: `rs3` **Command**: `rpush` **Key**: `"mylist"` **Args**: `'three`

Result field: `rs4` **Command**: `lset` **Key**: `"mylist"` **Args**: `[0,'four']`

Result field: `rs5` **Command**: `lset` **Key**: `"mylist"` **Args**: `[-2,'five']`

Result field: `rs6` **Command**: `lrange` **Key**: `"mylist"` **Args**: `[0,-1]`

Try This at Home

The Pipeline below contains only this example Function. You can import it into your own Cribl Edge environment, fill in the `url` with your own credentials, and further modify it to meet your needs.

redis-multiple-args.json

```
{
  "id": "redis-multiple-args",
  "conf": {
    "output": "default",
    "groups": {},
    "asyncFuncTimeout": 1000,
    "functions": [
      {
        "filter": "true",
        "conf": {
          "commands": [
            {
              "outField": "rs1",
              "command": "rpush",
              "keyExpr": "\"mylist\"",
              "argsExpr": "'one'"
            },
            {
              "outField": "rs2",
              "command": "rpush",
              "keyExpr": "\"mylist\"",
              "argsExpr": "'two'"
            },
            {
              "outField": "rs3",
              "command": "rpush",
              "keyExpr": "\"mylist\"",
              "argsExpr": "'three'"
            },
            {
              "command": "lset",
              "keyExpr": "\"mylist\"",
              "argsExpr": "[0,'four']",
              "outField": "rs4"
            },
            {
              "outField": "rs5",
              "command": "lset",
              "keyExpr": "\"mylist\"",
              "argsExpr": "[-2,'five']"
            },
            {
```

```
            "outField": "rs6",
            "command": "lrange",
            "keyExpr": "\"mylist\"",
            "argsExpr": "[0,-1]"
          }
        ],
        "maxBlockSecs": 60,
        "url": "redis://<your-credentials-here>"
      },
      "id": "redis",
      "disabled": false,
      "description": "Push arrays of multiple arguments to Redis"
    }
  ]
}
}
```

14.24. Regex Extract

The Regex Extract Function extracts fields using named capture groups. (In Splunk, these will be index-time fields). Fields that start with `__` (double underscore) are special in Cribl Edge. They are ephemeral: they can be used by any Function downstream, but **will not** be added to events, and **will not** exit the Pipeline.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of the Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Regex: Regex literal. Must contain named capturing groups, e.g.: `(?<foo>bar)`. Can contain special `_NAME_N` and `_VALUE_N` capturing groups, which extract **both the name and value** of a field, e.g.: `(?<_NAME_0>[^\s=]+)=(?<_VALUE_0>[^\s]+)`. Defaults to empty. See Examples below.

Additional regex: Click **Add Regex** to chain extra regex conditions.

Source field: Field on which to perform regex field extraction. Nested addressing is supported. Defaults to `_raw`.

Advanced Settings

Max exec: The maximum number of times to apply the **Regex** to the source field when the global flag is set, or when using `_NAME_N` and `_VALUE_N` capturing groups. Named capturing groups will always use a value of `1`. Defaults to `100`.

Field name format expression: JavaScript expression to format field names when `_NAME_n` and `_VALUE_n` capturing groups are used. E.g., to append `XX` to all field names, use: `` `${name}_XX` `` (backticks are literal). If not specified, names will be sanitized using regex: `/^[_0-9]+|[^a-zA-Z0-9_]+/g`. The **original** field name is in the global `name`. You can access other fields' values via `__e.<fieldName>`.

Overwrite existing fields: Whether to overwrite existing event fields with extracted values. If set to `No` (the default), existing fields will be converted to an array. If toggled to `Yes`, Regex Extract will create array fields if applied multiple times, or if fields exist. (E.g., if `src_ip` is extracted in an input Pipeline where it is assigned a value of `10.1.2.2`, and is also in a processing Pipeline with a value of `10.2.3.3`, then the resulting field is `["10.1.2.2", "10.2.3.3"]`.)

Examples

To test Regex Extract Functions use Data Preview to see the events as they flow into and out of the Pipeline.

Example 1: Single Field from Simple Event

Assume a simple event that looks like this: `metric1=23 metric2=42 dc=23 abc=xyz`

Extract **only** the `metric1` field:

Regex: `metric1=(?<metric1>\d+)` **Result**: `metric1:"23"`

Example 2: Extracting Multiple Fields

Use this sample data:

```
462559d4a487[471]: 172.23.0.6 - - [26/Feb/2024:20:22:38 +0000] "GET /catalog/view/
```

Use a regex to extract the fields you know are in the event.

Regex: `^(?<container_id>[^\s]+)\[(?<process_id>\d+)\]:\s+(?<remote_host>[^\s]+)\s+(?<remote_user>-)\s+(?<auth_user>-)\s+\[(?<timestamp>[^\]]+)\]\s+"(?<request_method>\w+)\s+(?<requested_url>[^\s]+)\s+(?<http_version>[^"]+)"\s+(?<status>\d+)\s+(?<bytes>.+)$`

Results:

Example 2 results

Example 3: Key-Value Pairs from Multiple Fields

Use this sample data:

```
rec_type=71 rec_type_simple=RNA dest_port=443 snmp_out=0 netflow_src="00000000-0000
```

Use a regex to extract **all** k=v pairs, then use **Field Name Format Expression** to append an `_XX` suffix to each extracted field:

Regex: `(?<_NAME_0>[\w-]+)="?(?<_VALUE_0>(?<=")[^"]*|\S*)` **Field Name Format Expression:** `${name}_XX`

Results:

Example 3 results

Example 4: Multi-Stage Extraction, Complex Events

This example builds on the syntax in Example 3, to tackle a more complex event structure.

In the right **Sample Data** pane, click **Paste** and insert the following sample:

Sample Data

```
<134>1 2020-12-22T17:06:08Z CORP_INT_NLB CheckPoint 18160 - [action:"Accept"; conn
```

This event is from a CheckPoint Firewall CMA system. With this type of event structure, properly extracting each event field into a separate metadata field requires two-stage processing. So we'll use two Regex Extract Functions.

The first Regex Function splits the event to separate the actual data from the header information. We'll split after the `CheckPoint 18160` string, by capturing everything between the `[` and `]`:

Regex: `\[(?<__fields>.*)\]` **Source**: `_raw`

Next, add this second Regex Extract Function to extract all k=v pairs:

Regex: `(?<_NAME_0>[^ :]+):(?<_VALUE_0>[^;]+);` **Source**: `__fields`

Results:

Example 4 results

For further examples, see Using Cribl to Analyze DNS Logs in Real Time – Part 2.

14.25. Regex Filter

The Regex Filter Function filters out events based on regex matches.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Regex: Regex to test against. Defaults to empty.

Additional regex: Click **Add Regex** to chain extra regex conditions.

Field: Name of the field to test against the regex. Defaults to `_raw`. Supports nested addressing.

Examples

See Regex Filtering for examples.

14.26. Rename

The Rename Function is designed to change fields' names or reformat their names (e.g., by normalizing names to camelcase). You can use Rename to change specified fields (much like the Eval Function), or for bulk renaming based on a JavaScript expression (much like the Parser Function).

Compared to these alternatives, Rename offers a streamlined way to alter only field names, without other effects.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Optionally, enter a simple description of this step in the Pipeline. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Parent fields: Specify fields whose children will inherit the **Rename fields** and **Rename expression** operations. Supports wildcards. If empty, only top-level fields will be renamed.

This Function cannot operate on internal fields whose names begin with double underscores (`__`).

Note that you can still use `__e` to access fields within the `(context)` object, as long as those fields do not begin with double underscores. That's because `__e` is a special variable, not a field.

Rename fields: Each row here is a key-value pair that defines how to rename fields. The current name is the key, and the new name is the value. Click **Add Field** to add more rows.

- **Current name**: Original name of the field to rename. You must quote literal identifiers (non-alphanumeric characters such as spaces or hyphens).
- **New name**: New or reformatted name for the field. Here again, you must quote literals.

Rename expression: Optional JavaScript expression whose returned value will be used to rename fields. Use the `name` and `value` global variables to access fields' names/values. Example: `name.startsWith('data') ? name.toUpperCase() : name`. You can access other fields' values via `event.<fieldName>`.

A single Function can include both **Rename fields** (to rename specified field names) and **Rename expression** (to globally rename fields). However, the **Rename fields** strategy will execute first.

Advanced Settings

Parent field wildcard depth: For wildcards specified in **Parent fields**, sets the maximum depth within events to match and rename fields. Enter `0` to match only top-level fields. Defaults to `5` levels down.

Example

Change the `level` field, and all fields that start with `out`, to all-uppercase.

Example event:

```
{"inEvents": 622,
  "level": "info",
  "outEvents": 311,
  "outBytes": 144030,
  "activeCxn": 0,
  "openCxn": 0,
  "closeCxn": 0,
  "activeEP": 105,
  "blockedEP": 0
}
```

Rename Fields:

Current name: `level`
New name: `LEVEL` **Rename expression**: `name.startsWith('out') ? name.toUpperCase() : name`

Event after Rename:

```
{"inEvents": 622,
  "LEVEL": "info",
  "OUTEVENTS": 311,
  "OUTBYTES": 144030,
  "activeCxn": 0,
  "openCxn": 0,
  "closeCxn": 0,
  "activeEP": 105,
  "blockedEP": 0
}
```

Applications

1. Remove filename prefix `<myPrefix>`:

Rename expression: `name.replace(/<myPrefix>/, '')`

2. Add a wildcard to rename a set of fields named `json.record[0]`, `json.record[1]`, etc., preserving the variable numbers in the brackets:

Rename expression: `name.replace(/(json)\.(\w+)/,'MYNEWNAME-$2-$1')`

14.27. Rollup Metrics

The Rollup Metrics Function merges/rolls up frequently generated incoming metrics into more manageable time windows.

Each Worker Process executes this Function independently on its share of events. For details, see Functions and Shared-Nothing Architecture.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Optional description of this Function's purpose in this Pipeline. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Dimensions: List of data dimensions across which to perform rollups. Supports wildcards. Defaults to `*` wildcard, meaning all original dimensions.

Time window: The time span over which to roll up (aggregate) metrics. Must be a valid time string (e.g., `10s`). Must match pattern: `\d+[sm]$`.

With high-cardinality data, beware of setting long time windows. Doing can cause high memory consumption and/or lost data, because memory is flushed upon restarts and redeployments.

Gauge update: The operation to use when rolling up gauge metrics. Defaults to **Last**; other options are **Maximum**, **Minimum**, or **Average**.

Examples

Scenario A:

Assume that you have metrics coming in at a rate that is too high. For example, Cribl Edge's internal metrics come in at a 2s interval.

To roll up these metrics to 1-minute granularity, you would set up the Rollup Metrics Function with a **Time Window** value of `60s`.

Scenario B:

Assume that you have metrics coming up with multiple dimensions – e.g. `host`, `source`, `data_center`, and `application`. You want to aggregate these metrics to eliminate some dimensions.

Here, you would configure Rollup Metrics Function with a **Time Window** value that matches the metrics' generation – e.g., `10s`. In the **Dimensions** field, you would remove the default `*` wildcard, and would specify only the dimensions you want to keep – e.g.: `host`, `data_center`.

14.28. Sampling

The Sampling Function filters out events, based on an expression and a sampling rate.

Each Worker Process executes this Function independently on its share of events. For details, see Functions and Shared-Nothing Architecture.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Sampling rules: Events matching these rules will be sampled at the rates you specify:

- **Filter**: Filter expression matching events to be sampled. Use `true` to match all.
- **Sampling rate**: Enter an integer `N`. (Defaults to `1`.) Sampling will pick `1/N` events matching this rule.

How It Works

Setting this Function's **Sampling rate** to `30` would mean that only 1 of every 30 events would be kept.

Let's assume that we save this setting, and then capture data from a datagen Source by selecting **Preview** > **Start a Capture** > **Capture**. In the **Capture Sample Data** modal, select: `100` seconds, `100` events, and **As they come in**. Then start the capture, and **Save as Sample File**.

Next, in the **Preview** pane, click **Simple** beside the new file's name. If you then click the **Basic Statistics** (chart) button, you should see that we've kept about 4 of the original 100 events, or close to 1 in 30.

	Full Event Length ⓘ	Number of Fields ⓘ	Number of Events ⓘ
IN	28.82KB	41	100
OUT	1.42KB	38	4
DIFF	↓ -95.08%	↓ -7.32%	↓ -96.00%

Examples

See Sampling for examples.

14.29. Serialize

Use the Serialize Function to serialize an event's content into a predefined format.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Type: Data output format. Defaults to `CSV`. Options:

- **CSV**
- **Extended Log File Format** – Extended Log File Format.
- **Common Log Format** – Common Log Format.
- **Key=Value Pairs** – selecting this type displays additional options:
  - **Clean Fields**: Toggle to `Yes` to clean field names by replacing non-alphanumeric characters with `_`. This will also strip leading and trailing `"` symbols.
  - **Pair delimiter**: Delimiter used to separate `key=value` pairs. Defaults to a single space character. Multiple characters are allowed. Cannot be an equal sign (`=`).
- **JSON Object**
- **Delimited values** – selecting this type displays additional options:
  - **Delimiter**: Delimiter character to split value. If left blank, defaults to comma (`,`). You can also specify pipe (`|`) or tab characters.
  - **Quote char**: Character used to quote literal values. If left blank, defaults to `"`.
  - **Escape char**: Character used to escape delimiter or quote characters. If left blank, defaults to **Quote char**.
  - **Null value**: Field value representing the null value. These fields will be omitted. Defaults to: `-`

Library: Browse Parser/Formatter library.

Fields to serialize: You must specify individual fields with the `CSV`, `ELFF`, `CLF`, and `Delimited values` Types. All other formats support wildcard field lists. With these formats, the default entry here will specify some excluded fields (indicated by `!`), followed by a `*` wildcard to serialize all remaining fields.

All field names or wildcards that you want to exclude from serialization must precede all field names or wildcards that you want to include.

Source field: Field containing the object to serialize. Leave blank to serialize top-level event fields.

Destination field: Field to serialize the data into. Defaults to `_raw`.

Examples

Scenario A: JSON to CSV

Assume a simple event that looks like this: `{"time":"2019-08-25T14:19:10.240Z","channel":"input","level":"info","message":"initializing input","type":"kafka"}`

We want to serialize these fields: `_time`, `channel`, `level`, and `type` into a single string, in CSV format, stored in a new destination field called `test`.

To properly extract the key-value pairs from this event structure, we'll use a built-in Event Breaker:

1. Copy the above sample event to your clipboard.
2. In the **Preview** pane, select **Paste a Sample**, and paste in the sample event.
3. Under **Select Event Breaker**, choose **ndjson** (newline-delimited JSON), and click **Save as a Sample File**.

Now you're ready to configure the Serialize Function, using the settings below:

Type: CSV Fields to Serialize: `_time channel level type` Destination Field: `test` Source Field: [leave empty] **Result**: `test: 1566742750.24,input,info,kafka`

In the new `test` field, you now see the `time`, `channel`, `level`, and `type` keys extracted as top-level fields.

Scenario B: CSV to JSON

Let's assume that a merchant wants to extract a subset of each customer order, to aggregate anonymized order statistics across their customer base. The transaction data is originally in CSV format, but the statistical data must be in JSON.

Here's a CSV header (which we don't want to process), followed by a row that represents one order:

```
orderID,custName,street,city,state,zip 20200622102822,john smith,100 Main St.,Anytown,AK,99911
```

To convert to JSON, we'll need to first parse each field from the CSV to a manipulable field in the Pipeline, which the Serialize Function will be able to reference. In this example, the new manipulable field is `message`.

Use the `Parser` Function:

Filter: `true` Operation mode: `Extract` Type: CSV Source field: `_raw` Destination field: `message` List of fields: `orderID custName street city state zip`

Now use the Serialize Function:

Filter: `true` Type: `JSON` Fields to serialize: `city state` Source field: `message` Destination field: `orderStats`

14.30. Suppress

The Suppress Function suppresses events over a time period, based on evaluating a key expression.

Each Worker Process executes this Function independently on its share of events. For details, see Functions and Shared-Nothing Architecture.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Key expression: Suppression key expression used to uniquely identify events to suppress. For example, `` `${ip}:${port}` `` will use the fields `ip` and `port` from each event to generate the key.

Number to allow: The number of events to allow per time period. Defaults to `1`.

Suppression period (sec): The number of seconds to suppress events after 'Number to allow' events are received. Defaults to `300`.

Drop suppressed events: Specifies if suppressed events should be dropped, or just tagged with `suppress=1`. Defaults to `Yes`, meaning drop.

Advanced Settings

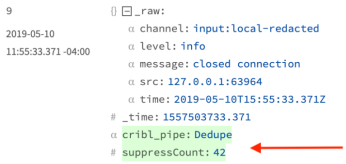
Maximum cache size: The maximum number of keys that can be cached before idle entries are removed. Before changing the default `50000`, contact Cribl Support to understand the implications.

Suppression period timeout: The number of suppression periods of inactivity before a cache entry is considered idle. This defines a multiple of the **Suppression period (sec)** value. Before changing the default `2`, contact Cribl Support to understand the implications.

Num events to trigger cache clean-up: Check cache for idle sessions every N events when cache size exceeds the **Maximum cache size**. Before changing the default `10000`, contact Cribl Support to understand the implications.

Seeing the Results

If you've enabled **Drop suppressed events**, such events will be omitted from logs as they exit this Function. However, the next event allowed through will include a `suppressCount: N` field, whose `N` value indicates the number of events dropped in the preceding **Suppression period**.



suppressCount shows number of events dropped

Examples

In the examples below, **Filter** is the Function-level Filter expression:

1. Suppress by the value of the `host` field: Filter: `true` Key expression: `host` Number to allow: `1` Suppression period (sec): `30`
   Using a datagen sample as a source, generate at least 100 events over 2 minutes.

Result: One event per unique `host` value will be allowed in every 30s. Events without a `host` field will **not** be suppressed.

2. Suppress by the value of the `host` and `port` tuple : Filter: `true` Key expression: `` `${host}:${port}` `` Number to allow: `1` Suppression period (sec): `300`

Result: One event per unique `host:port` tuple value will be allowed in every 300s.

Suppression will **also** apply to events without a `host` or a `port` field. The reason is that if `field` is not present, `` `${field}` `` results in the literal `undefined`.

3. To **guarantee** that suppression applies **only** to events with `host` and `port`, check for their presence using a Filter: Filter: `host && port!=null` Key expression: `` `${host}:${port}` `` Number to allow: `1` Suppression period (sec): `300`
4. Decorate events that qualify for suppression: Filter: `true` Key expression: `` `${host}:${port}` `` Number to allow: `1` Suppression period (sec): `300` Drop suppressed events: `No`

Result: No events will be suppressed. But all qualifying events will gain an added field `suppress=1`, which can be used downstream to further transform these events.

For further use cases, see Cribl's Streaming Data Deduplication with Cribl blog post. For further details on filter usage, see Filters.

14.31. Tee

The Tee Function tees events out to a command of choice, via `stdin`. The output is one JSON-formatted event per line. You can send the events to (for example) a local file on the Cribl Edge worker. This can be useful in verifying the data being processed in a Pipeline.

The Filesystem/NFS Destination offers similar capability, but only after the data leaves the Pipeline. Tee, by comparison, can be inserted at any point in the Pipeline.

In Cribl.Cloud, the Tee Function is only available on hybrid, customer-managed Edge Nodes.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Command: Command to execute and receive events (via `stdin`) – one JSON-formatted event per line.

Args: Click **Add Arg** to supply arguments to the command.

Restart on exit: Restart the process if it exits and/or we fail to write to it. Defaults to `Yes`.

Environment variables: Environment variables to set or overwrite. Click **Add Variable** to add key-value pairs.

Communication Protocol

Data is passed to the command through its `stdin`, using the following protocol:

- First line: Metadata serialized in JSON, containing the following fields:
  - **format**: Serialization format for event. Defaults to `JSON`.
  - **conf**: Full Function configuration.
- Remaining: Payload.

Examples

Assume that we are parsing PANOS Traffic logs, and want to see how they look at a particular step in the processing Pipeline We'll assume that the `Parser` Function is already in place, so we'll insert the Tee Function at any (arbitrary) later point in the Pipeline.

Scenario A:

The Tee Function itself requires only that we define the **Command** field. In this particular example, that **Command** will be `tee` itself.

We've also clicked **Add Arg**, to specify a local output file in the resulting **Args** field. (A file path would normally be the first argument to a `tee` command executed from the command line. The Cribl Edge user must have write permission on the specified file path.)

Command: `tee`

Args: `/opt/cribl/foo.log`

In this first scenario, assume that we have the `Parser` configured to parse, but not keep any fields. After changes are deployed and PANOS logs are received, if we tail `foo.log`, we'd see the following:

```
Line 1: {"format":"json","conf":{"restartOnExit":true,"env":
{},"command":"tee","args":["/opt/cribl/foo.log"]}

Line 2: {"_raw":"Oct 09 10:19:15 DMZ-internal.nsa.gov 1,2019/10/09
10:19:15,001234567890002,TRAFFIC,drop,2304,2019/10/09
10:19:15,209.118.103.150,160.177.222.249,0.0.0.0,0.0.0.0,InternalServer,,,not-
applicable,vsys1,inside,z1-FW-Transit,ethernet1/2,,All traffic,2019/10/09
10:19:15,0,1,63712,443,0,0,0x0,udp,deny,60,60,0,1,2019/10/09
10:19:15,0,any,0,0123456789,0x0,Netherlands,10.0.0.0-10.255.255.255,0,1,0,policy-
deny,0,0,0,0,,DMZ-internal,from-policy,,,0,,0,,N/A,0,0,0,0,1202585d-b4d5-5b4c-aaa2-
d80d77ba456e,0","_time":1593185574.663,"host":"127.0.0.1"}
```

In Line 2 above, note that the `_raw` field makes up most of the contents, with only the `_time` and `host` fields added.

Scenario B:

Assume that we use the Tee Function, using the same **Command** and arguments, but we've modified the `Parser` Function to retain five fields: `receive_time`, `source_port`, `destination_port` `bytes_received`, and `packets_received`.

This time, if we tail `foo.log`, we'll see something like the following. If you compare this output to the previous output example, you'll notice the five fields appended to this event:

```
Line 3: {"_raw":"Oct 09 10:19:15 DMZ-internal.nsa.gov 1,2019/10/09
10:19:15,001234567890002,TRAFFIC,drop,2304,2019/10/09
10:19:15,209.118.103.150,160.177.222.249,0.0.0.0,0.0.0.0,InternalServer,,,not-
applicable,vsys1,inside,z1-FW-Transit,ethernet1/2,,All traffic,2019/10/09
10:19:15,0,1,63712,443,0,0,0x0,udp,deny,60,60,0,1,2019/10/09
10:19:15,0,any,0,0123456789,0x0,Netherlands,10.0.0.0-10.255.255.255,0,1,0,policy-
deny,0,0,0,0,,DMZ-internal,from-policy,,,0,,0,,N/A,0,0,0,0,1202585d-b4d5-5b4c-aaa2-
d80d77ba456e,0","_time":1593185606.965,"host":"127.0.0.1","receive_time":"2019/10/09
10:19:15","source_port":"63712","destination_port":"443","bytes_received":"0","packets
```

In this Function's **Command** field, you can specify commands other than `tee` itself. For example: By using `nc` as the command, and specifying `localhost` and a port number (as two separate arguments), you'll see event data being received via `nc` on the specified port.

14.32. Trim Timestamp

The Trim Timestamp Function removes timestamp patterns from events, and (optionally) stores them in a specified field.

This Function looks for a timestamp pattern that exists between the characters indicated by numeric `timestartpos` and `timeendpos` fields. It removes `timestartpos` and `timeendpos` along with the timestamp pattern.

The Trim Timestamp Function, in current Cribl Edge versions, removes timestamps only from events whose `timestartpos` value is set to `0`. If you need to strip timestamps from arbitrary postions within events, instead use an Eval Function.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description about this step in the Pipeline. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Field name: Name of field in which to save the timestamp. (If empty, timestamp will not be saved to a field.)

Example

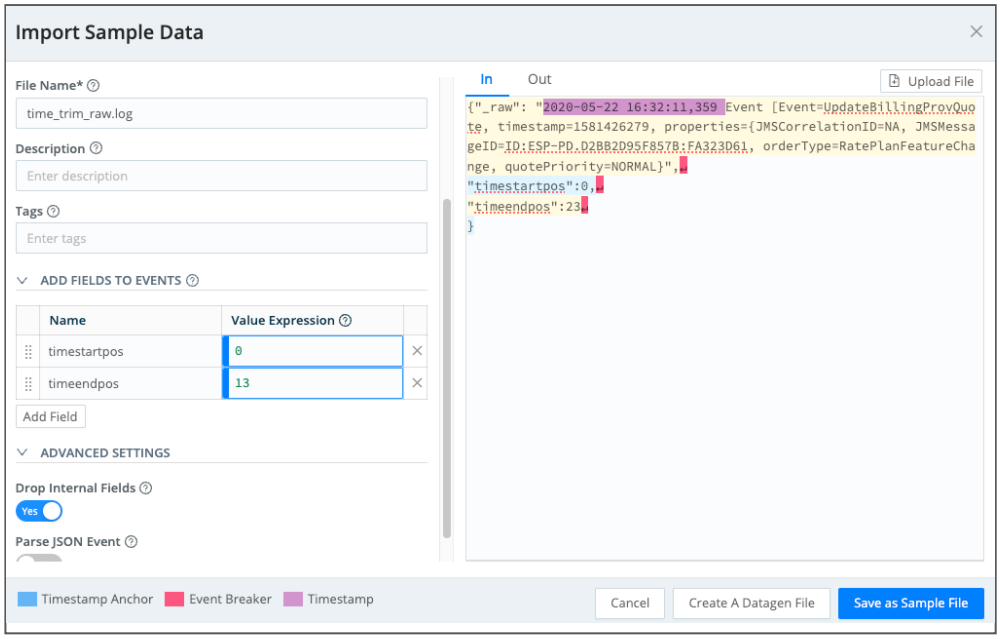
Remove the timestamp pattern (indicated by `timestartpos` and `timeendpos`) from `_raw`, and stash it in a field called `time_field`.

Field name: `time_field`

Example event before:

```
{"_raw": "2020-05-22 16:32:11,359 Event [Event=UpdateBillingProvQuote, timestamp=1
"timestartpos":0,
"timeendpos":23
}
```

To create this example payload, we selected **Sample Data** > **Import**, pasted the `_raw` field's original contents into the resulting modal, and then added the two required position fields:

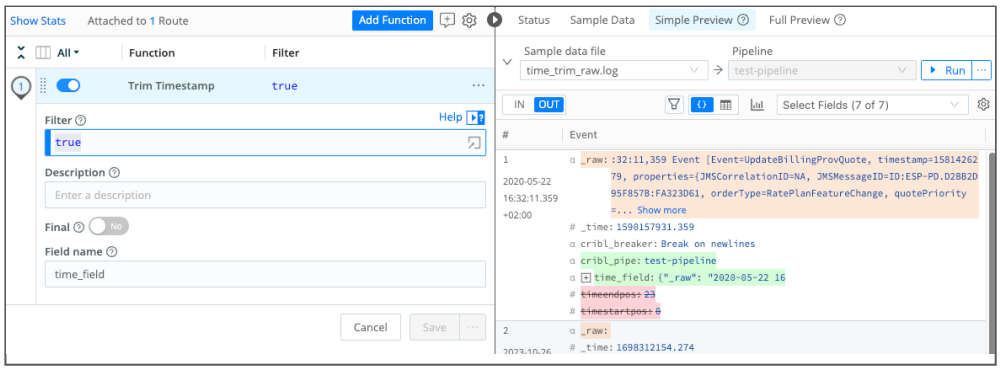


Example event setup

Example event after:

```
{"_raw": "Event [Event=UpdateBillingProvQuote, timestamp=1581426279, properties={JM
"time_field":"2020-05-22 16:32:11,359"
}
```

In the Preview pane's **OUT** view, the original timestamp has been removed from `_raw`, and lifted into the new `time_field` we specified in the Function. The `timestartpos` and `timeendpos` fields have been removed.



Example event, saved and transformed

14.33. Unroll

The Unroll Function accepts an array field – or an expression to evaluate an array field – and breaks/unrolls the array into individual events.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Source field expression: Field in which to find/calculate the array to unroll. E.g.: `_raw`, `_raw.split(/\n/)`. Defaults to `_raw`.

Destination field: Field (within the destination event) in which to place the unrolled value. Defaults to `_raw`.

Example

Assume we want to break/unroll each line of this event:

Sample Event

```
USER       PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
root         1  0.0  0.5  38000  5356 ?        Ss    2018   2:02 /lib/systemd/syste
root         2  0.0  0.0      0     0 ?        S     2018   0:00 [kthreadd]
root         3  0.0  0.0      0     0 ?        S     2018   1:51 [ksoftirqd/0]
root         5  0.0  0.0      0     0 ?        S<    2018   0:00 [kworker/0:0H]
root         7  0.0  0.0      0     0 ?        S     2018   3:55 [rcu_sched]
root         8  0.0  0.0      0     0 ?        S     2018   0:00 [rcu_bh]
```

Settings

Source field expression: `_raw.split(/\n/)`

The `split()` JavaScript method breaks `_raw` into an ordered set of substrings/values, puts these values into an array, and returns the array.

Destination field: `_raw`

Resulting Events

```
Event 1:
USER       PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND

Event 2:
root         1  0.0  0.5  38000  5356 ?        Ss    2018   2:02 /lib/systemd/syste

Event 3:
root         2  0.0  0.0      0     0 ?        S     2018   0:00 [kthreadd]

Event 4:
root         3  0.0  0.0      0     0 ?        S     2018   1:51 [ksoftirqd/0]

Event 5:
root         5  0.0  0.0      0     0 ?        S<    2018   0:00 [kworker/0:0H]

Event 6:
root         7  0.0  0.0      0     0 ?        S     2018   3:55 [rcu_sched]

Event 7:
root         8  0.0  0.0      0     0 ?        S     2018   0:00 [rcu_bh]
```

14.34. XML Unroll

The XML Unroll Function accepts a proper XML event with a set of elements, and converts the elements into individual events.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Unroll elements regex: Path to the array to unroll. E.g.: `^root\.child\.ElementToUnroll$`

Copy elements regex: Regex matching elements to copy into each unrolled event. E.g.: `^root\.(childA|childB|childC)$`

Unroll index field: Cribl Edge will add a field with this name, containing the 0-based index at which the element was located within the event. In Splunk, this will be an index-time field. Supports nested addressing. Name defaults to `unroll_idx`.

Pretty print: Whether to pretty print the output XML.

Examples

Assume that the following sample is ingested as a single event:

sample.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<Parent>
    <myID>123456</myID>
    <branchLocation>US</branchLocation>
    <Child>
        <state>NY</state>
        <city>New York</city>
    </Child>
    <Child>
        <state>NJ</state>
        <city>Edgewater</city>
    </Child>
    <Child>
        <state>CA</state>
        <city>Oakland</city>
    </Child>
    <Child>
        <state>CA</state>
        <city>San Francisco</city>
    </Child>
</Parent>
```

If you insert this sample using **Preview** > **Add a Sample** > **Paste a Sample**, adjust Event Breaker settings to add the sample as a single event. One way to do this is to add a regex Event Breaker that (by design) will not match anything present in the sample. For example: `/[\n\r]+donotbreak(?!\s)/`. In current Cribl Edge versions, you can also use the built-in **Do Not Break** Ruleset.

Set up the XML Unroll Function using these settings:

Unroll elements regex: `^Parent\.Child$` **Copy elements regex**: `^Parent\.(myID|branchLocation)$`

Output 4 Events:

Resulting Events

```
# Event 1
<?xml version="1.0"?>
<Child>
  <myID>123456</myID>
  <branchLocation>US</branchLocation>
  <state>NY</state>
  <city>New York</city>
</Child>

# Event 2
<?xml version="1.0"?>
<Child>
  <myID>123456</myID>
  <branchLocation>US</branchLocation>
  <state>NJ</state>
  <city>Edgewater</city>
</Child>

# Event 3
<?xml version="1.0"?>
<Child>
  <myID>123456</myID>
  <branchLocation>US</branchLocation>
  <state>CA</state>
  <city>Oakland</city>
</Child>

# Event 4
<?xml version="1.0"?>
<Child>
  <myID>123456</myID>
  <branchLocation>US</branchLocation>
  <state>CA</state>
  <city>San Francisco</city>
</Child>
```

14.35. Prometheus Publisher (Deprecated)

This Function was deprecated as of Cribl Edge 3.0, and has been removed from Cribl Edge as of v.4.0. Please instead use the Prometheus Destination to send metrics to Prometheus-compatible endpoints.

The Prometheus Publisher Function allows for metrics to be published to a Prometheus-compatible metrics endpoint. These can be upstream metrics received by Cribl Edge, or metrics derived from the output of Cribl Edge's `Publish Metrics` or `Aggregation` Functions. A Prometheus instance is responsible for collecting the metrics at that endpoint, and for performing its own processing of the metric data.

In the current Cribl Edge version, the endpoint is: `http://<worker_node_IP>:<api-port>/metrics`. Within Cribl Edge, that endpoint redirects from `http://<worker_node_IP>:9000/metrics` to `http://<worker_node_IP>:9000/api/v1/metrics`.

If used, this Function **must** follow any Publish Metrics or Aggregations Functions within the same Pipeline. This is to ensure that any data **not** originating from a metrics input is transformed into metrics format.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Fields to publish: Wildcard list of fields to publish to the Prometheus endpoint.

Advanced Settings

Batch write interval: How often, in milliseconds, the contents should be published. Defaults to `5000`.

Passthrough mode: If set to `No` (the default), overrides the **Final** setting, and suppresses output to downstream Functions' Destinations. Toggle to `Yes` to allow events to flow to consumers beyond the Prometheus endpoint. In effect, when previewing the pipeline output what you'll see is your event fields will have strikethrough font applied to them. This does not mean the Prometheus function is not matching your events but rather indicative of the Passthrough being disabled.

Update mode: On the default `No` setting, suppresses output to downstream Functions' Destinations. (This overrides the **Final** setting.) Toggle to `Yes` to allow events to flow to consumers beyond the Prometheus endpoint.

Example

This example uses the same PANOS sample data as the Publish Metrics Function, and is similarly preceded in a Pipeline by a Parser Function that extracts fields from the PANOS log.

Filter: Set as appropriate. **Fields to publish**: Set as appropriate. We'll use the default of `*` for this example. **Advanced settings**: Accept defaults.

After committing and deploying changes, you should be able to use a `curl` command (-L needed to follow the redirect mentioned above) to verify that metrics are being published, just a few seconds after data is ingested on an idle system.

curl output

```
$ curl -L http://<worker_node_IP>:9000/metrics
# TYPE perf_192_168_1_248_bytes_sent counter
metric_192_168_1_248_bytes_sent {destination_ip="160.177.222.249",inbound_interface

# TYPE perf_192_168_1_248_bytes_rcvd counter
metric_192_168_1_248_bytes_rcvd {destination_ip="160.177.222.249",inbound_interface

# TYPE perf_192_168_1_248_pkts_sent counter
metric_192_168_1_248_pkts_sent {destination_ip="160.177.222.249",inbound_interface=

# TYPE perf_192_168_1_248_pkts_rcvd counter
metric_192_168_1_248_pkts_rcvd {destination_ip="160.177.222.249",inbound_interface=
```

Now, we need to have Prometheus scrape the metrics. In this very basic example, you can add the target endpoint to the `prometheus.yml` file, under the `scrape_configs` -> `static_configs` section. Specify the endpoint in `IP:port` syntax, because Prometheus assumes (and requires) `/metrics` for all endpoints.

Restart Prometheus. Within just a few seconds, you should be able to use its query interface to retrieve metrics published by Cribl Edge.

14.36. Reverse DNS (Deprecated)

The Reverse DNS Function resolves hostnames from a numeric IP address, using a reverse DNS lookup.

This Function was deprecated as of v.2.4, and has been removed from Cribl Edge as of v.4.0. Please instead use the DNS Lookup Function's reverse lookup feature.

Usage

Filter: Filter expression (JS) that selects data to feed through the Function. Defaults to `true`, meaning it evaluates all events.

Description: Simple description of this Function. Defaults to empty.

Final: If toggled to `Yes`, stops feeding data to the downstream Functions. Defaults to `No`.

Lookup Fields

Lookup field name: Name of the field containing the IP address to look up.

If the field value is not in IPv4 or IPv6 format, the lookup is skipped.

Output field name: Name of the field in which to add the resolved hostname. Leave blank to overwrite the lookup field.

Reload period (minutes): How often to refresh the DNS cache. Use `0` to disable refreshes. Defaults to `60` minutes.

Example

Lookup field name: `dest_ip` **Output field name**: `dest_host` **Result**: See the `dest_ip` field, and the newly created `dest_host` field, in the events.


1
2020-06-29 12:51:03.551 -07:00
_raw: rec_type=71 rec_type_simple=RNA dest_port=443 snmp_out=0 netflow_src=00000000-0000-0000-0000-000000000000 ssl_server_cert_status="Not Checked" dest_ip=8.8.8.8 sec_intel_event=No mac_address=00:00:0... Show more
_time: 1593460263.551
app_proto: HTTPS
client_app: SSL client
client_version:
connection_id: 21378
cribl_breaker: Break on newlines
cribl_pipe: rev_dns
dest_autonomous_system: 0
dest_bytes: 3746
dest_host: dns.google
dest_ip: 8.8.8.8

15. Knowledge Libraries

Select **Processing** > **Knowledge** to access Cribl Edge's libraries of default Lookup tables, Event Breaker rulesets, Parsers, Global Variables, Regex expressions, Grok patterns, Schemas, Database Connections, and AppScope config templates. You can customize these libraries by adding your own Knowledge objects.

From the **Processing** > **Knowledge** top nav, you can select:

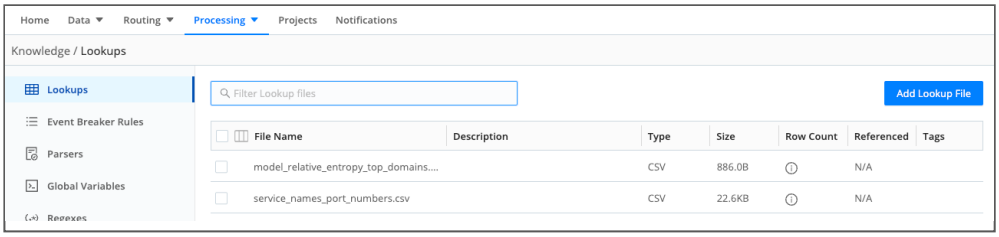
- Lookups
- Event Breakers
- Parsers
- Global Variables
- Regexes
- Grok Patterns
- Schemas (JSON)
- Parquet Schemas
- Database Connections (currently supported only in Cribl Stream)
- AppScope Configs

15.1. Lookups Library

What Are Lookups

Lookups are data tables that can be used in Cribl Edge to enrich events as they are processed by the Lookup Function. You can access the Lookups library, which provides a management interface for all lookups, from Cribl Edge's top nav under **Processing** > **Knowledge** > **Lookups**.

This library is searchable, and each lookup can be tagged as necessary. There's full support for `.csv` files. Compressed files are supported, but must be in gzip format (`.gz` extension). You can add files in multimedia database (`.mmdb`) binary format, but you cannot edit these binary files through Cribl Edge's UI.



Lookups Library

How Does the Library Work

In single-instance deployments, all files handled by the interface are stored in `$CRIBL_HOME/data/lookups`. In distributed deployments, the storage path on the Leader Node is `$CRIBL_HOME/groups/<groupname>/data/lookups/` for each Worker Group.

For large and/or frequently replicated lookup files, you might want to bypass the Lookups Library UI and instead manually place the files in a different location. This can both reduce deploy traffic and prevent errors with Cribl Edge's default Git integration. For details, see Managing Large Lookups.

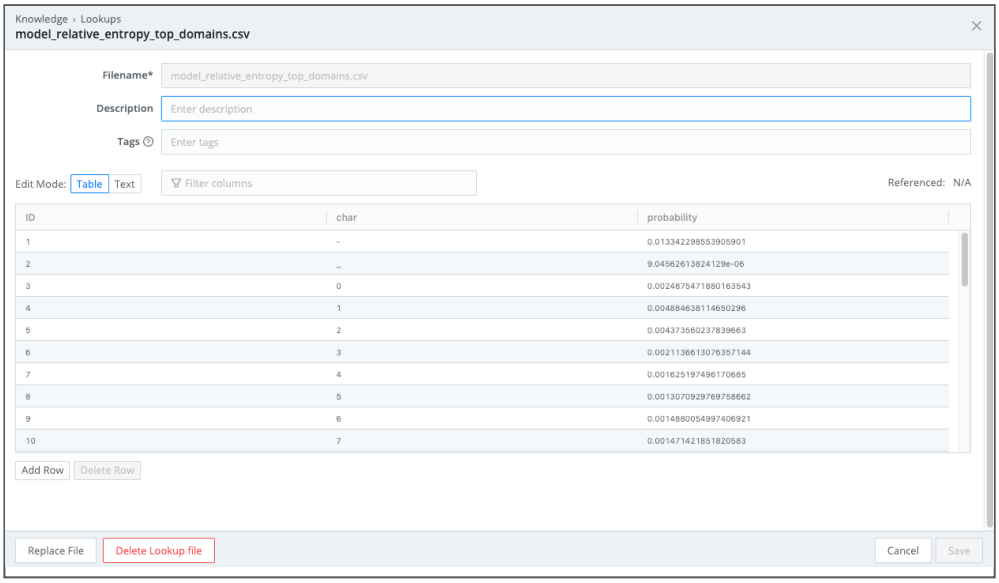
Adding Lookup Files

To upload or create a new lookup file, click **New Lookup File**, then click the appropriate option from the drop-down.

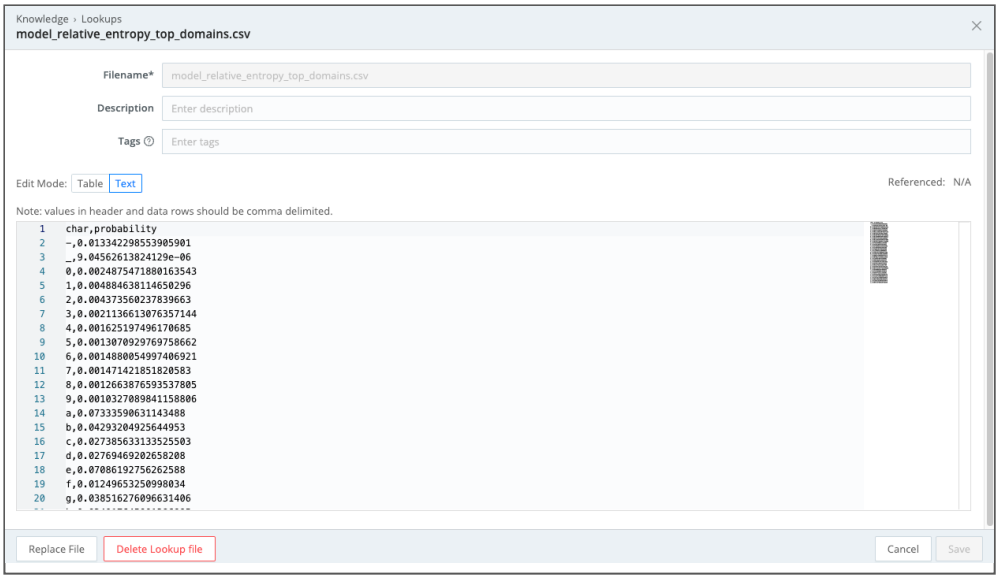
Modifying Lookup Files

To re-upload, expand, edit, or delete an existing `.csv` or `.gz` lookup file, click its row on the Lookups page. (No editing option is available for `.mmdb` files; you can only re-upload or delete these.)

In the resulting modal, you can edit files in **Table** or **Text** mode. However, **Text** mode is disabled for files larger than 1 MB.



Editing in **Table** mode



Editing in **Text** mode

Editing Regex with Quotation Marks

When a regex contains quotes, edit it in **Text** mode. Do not edit it **Table** mode, which can cause the regex to behave in unexpected ways.

Memory Sizing for Large Lookups

For large lookup files, you'll need to provide extra memory beyond basic requirements for Cribl Edge and the OS. To determine how much extra memory to add to a Worker Node/Edge Node for a lookup, use this formula:

```
Lookup file's uncompressed size (MB) * 2.25 (to 2.75) *  numWorkerProcesses = Extra RAM required for lookup
```

For example, if you have a lookup file that is 1 GB (1,000 MB) on disk, and three Worker Processes, you could use an average 2.50 as the multiplier:

```
1,000 * 2.50 * 3 = 7,500
```

In this case, the Node's server or VM would need an extra 7,500 MB (7.5 GB) to accommodate the lookup file across all three worker processes.

What's with That Multiplier?

We've cited a squishy range of 2.25–2.75 for the multiplier, because we've found that it varies inversely with the number of columns in the lookup file:

- The fewer columns, the higher the extra memory overhead (2.75 multiplier).
- The more columns, the lower the overhead (2.25 multiplier).

In Cribl's testing:

- 5 columns required a multiplier of 2.75
- 10 columns required a multiplier of only 2.25.

These are general (not exact) guidelines, and this multiplier depends only on the lookup table's number of columns. The memory overhead imposed by each additional row appears to decline when more columns are present in the data.

Maximum Table Size

Aside from the memory requirements above, the Node.js runtime imposes a hard limit on the size of lookup tables that Cribl Edge can handle. No table can contain more than 16,777,216 (i.e., 2^{24}) rows. If a lookup exceeds this size, attempting to load the lookup will log errors of the form: `failed to load function...Value undefined out of range...`.

Other Scenarios

See also:

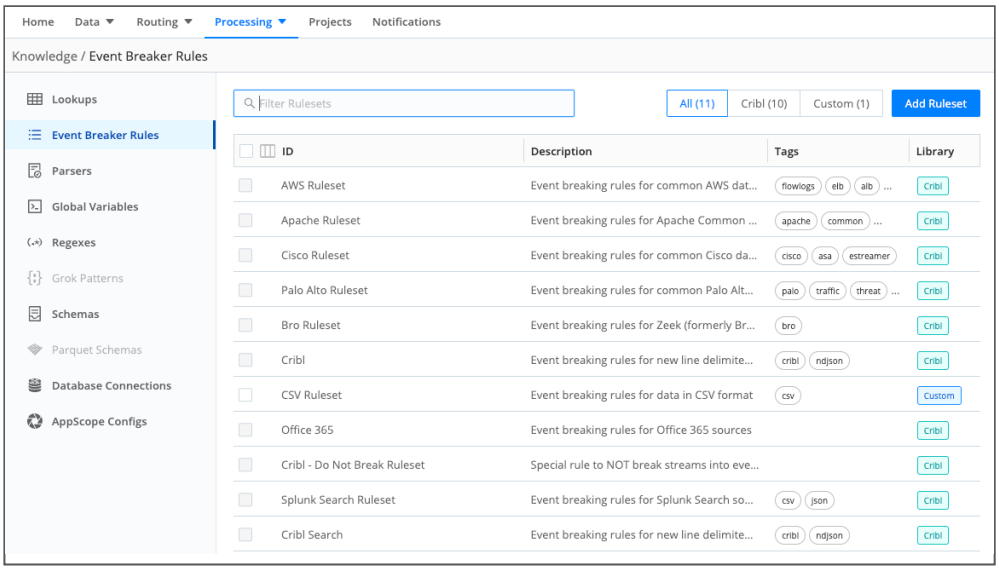
- Lookup Function.
- Ingest-time Lookups use case.
- Managing Large Lookups use case.
- Redis Function for faster lookups using a Redis integration (bypasses the Lookups Library).

15.2. Event Breakers

Event Breakers help break incoming streams of data into discrete events. To see how Event Breakers interact with the rest of Cribl Edge's data flow, see Event Processing Order.

Accessing Event Breakers

You access the Event Breakers management interface from Cribl Edge's top nav under **Processing** > **Knowledge** > **Event Breaker Rules**. On the resulting **Event Breaker Rules** page, you can edit, add, delete, search, and tag Event Breaker rules and rulesets, as necessary.



Event Breaker Rulesets page

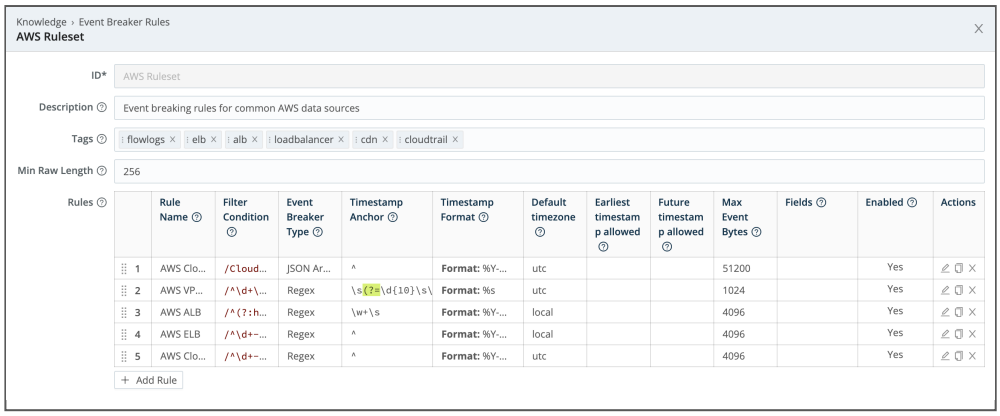
Limitations

Event Breakers are directly accessible only on Sources that require incoming events to be broken into a better-defined format. (Check individual Cribl Edge Sources' documentation for Event Breaker support.) Also, Event Breakers are currently not supported in Packs.

However, you can instead use the Event Breaker Function in Pipelines that process data from unsupported Sources, and in Pipelines within Packs.

Event Breaker Rulesets

Rulesets are **collections of Event Breaker rules** that are associated with Sources. Rules define configurations needed to break down a stream of data into events.



Rules within an example (AWS) ruleset that ships with Cribl Edge

Rules within a ruleset are ordered and evaluated top->down. One or more rulesets can be associated with a Source, and these rulesets are also evaluated top->down. For a stream from a given Source, the first matching rule goes into effect.

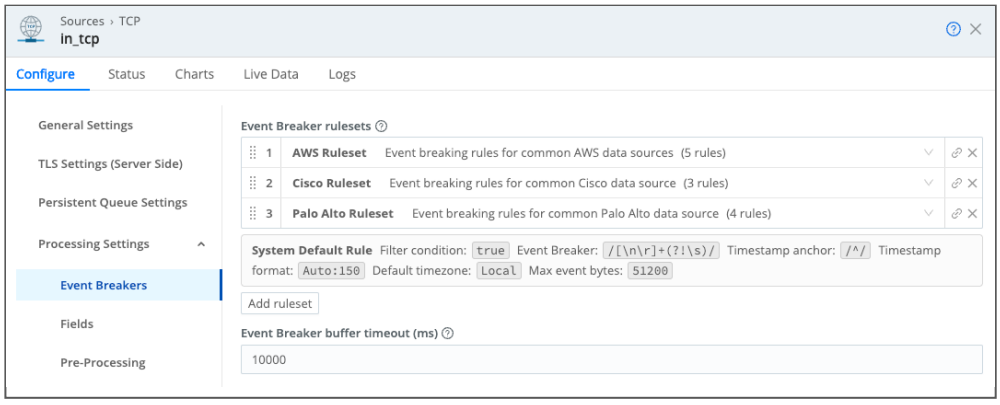
Rulesets and Rules - Ordered

```
Ruleset A
  Rule 1
  Rule 2
  ...
  Rule n

...

Ruleset B
  Rule Foo
  Rule Bar
  ...
  Rule FooBar
```

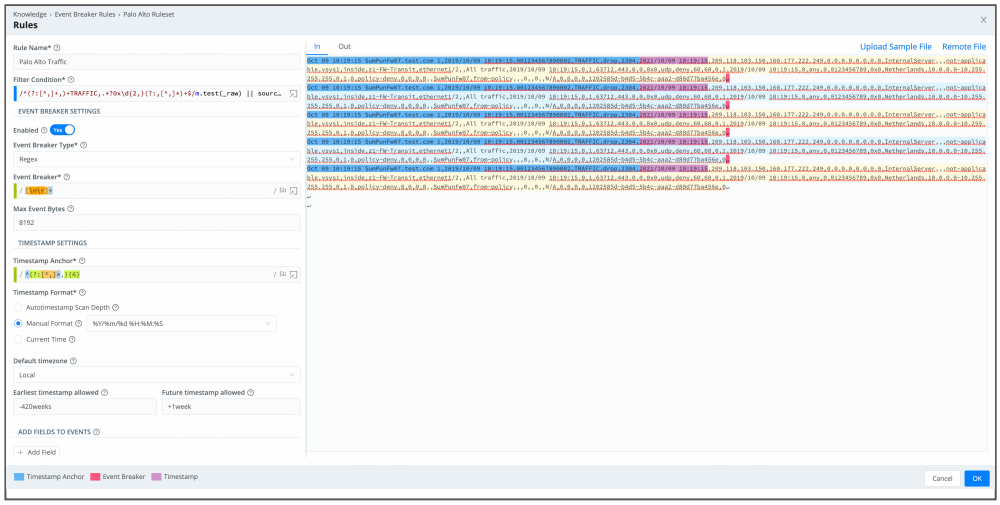
An example of multiple rulesets associated with a Source:



Three Event Breaker rulesets added to a Source

Rule Example

This rule breaks on newlines and uses Manual timestamping **after** the sixth comma, as indicated by this pattern: `^(?:[^,]*,){6}`.



An Event Breaker rule

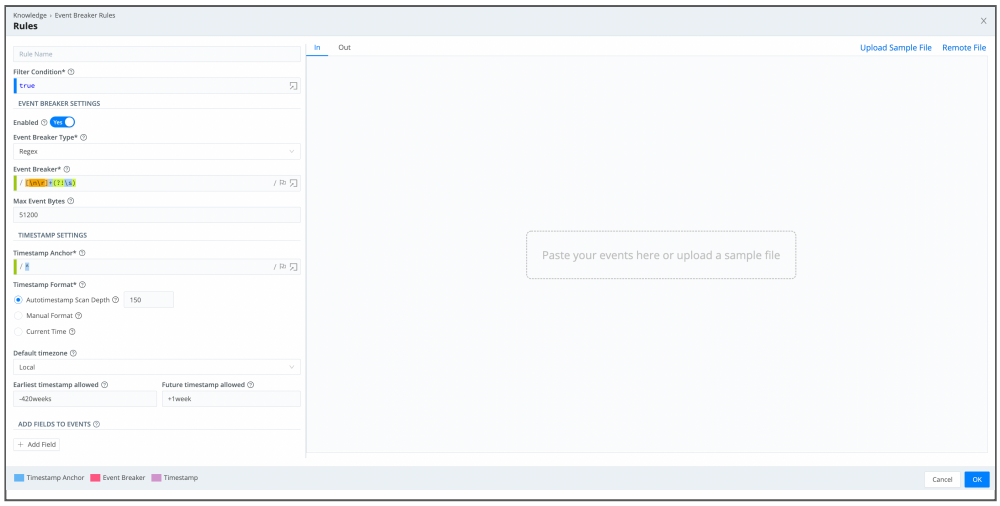
System Default Rule

The system default rule functionally sits at the bottom of the ruleset/rule hierarchy (but is built-in and not displayed on the Event Breakers page), and goes into effect if there are no matching rules:

- Filter Condition defaults to `true`
- Event Breaker to `[\n\r]+(?!\s)`
- Timestamp anchor to `^`
- Timestamp format to `Auto` and a scan depth of `150` bytes
- Max Event Bytes to `51200`
- Default Timezone to `Local`

How Do Event Breakers Work

On the **Event Breaker Rules** page (see screenshot above), click **Add Ruleset** to create a new Event Breaker ruleset. Click **Add Rule** within a ruleset to add a new Event Breaker.



Adding a new Event Breaker rule

Each Event Breaker includes the following components, which you configure from top to bottom in the above **Event Breaker Rules** modal:

Filter Condition

As a stream of data moves into the engine, a rule's filter expression is applied. If the expression evaluates to `true`, the rule configurations are engaged for the entire duration of that stream. Otherwise, the next rule down the line is evaluated.

The "stickiness" of the first `true` expression can lead to some unintended behavior. For example, if an S3 Source ingests an archive file that contains multiple files in different formats – where each file requires its own unique Event Breaker – events will not be broken correctly.

Event Breaker Type

After a breaker pattern has been selected, it will apply on the stream **continuously**. See below for specific information on different Event Breaker Types.

Timestamp Settings

After events are synthesized out of streams, Cribl Edge will attempt timestamping. First, a timestamp anchor will be located inside the event. Next, starting there, the engine will try to do one of the following:

- Scan up to a configurable depth into the event and autotimestamp, **or**
- Timestamp using a manually supplied `strptime` format, **or**
- Timestamp the event with the current time.

The closer an anchor is to the timestamp pattern, the better the performance and accuracy – especially if multiple timestamps exist within an event. For the manually supplied option, the anchor must lead the

engine **right before** the timestamp pattern begins.

Anchors preceding timestamps

This timestamping executes the same basic algorithm as the Auto Timestamp Function and the C.Time.timestampFinder() native method.

Cribl Edge truncates timestamps to three-digit (milliseconds) resolution, omitting trailing zeros.

In Cribl Edge 3.4.2 and above, where an Event Breaker has set an event's `_time` to the current time – rather than extracting the value from the event itself – it will mark this by adding the internal field `__timestampExtracted: false` to the event.

Add Fields to Events

After events have been timestamped, one or more fields can be added here as key-value pairs. In each field's **Value Expression**, you can fully evaluate the field value using JavaScript expressions.

Event Breakers **always** add the `cribl_breaker` field to output events. Its value is the name of the chosen ruleset. (Some examples below omit the `cribl_breaker` field for brevity, but in real life the field is always added.)

Max Event Bytes

If an event's final broken chunk reaches a matched rule's **Max event bytes** length, it will be broken again.

The highest **Max Event Bytes** value that you can set is about 128 MB (`134217728` bytes). Events exceeding that size will be split into separate events, but left unbroken. Cribl Edge will set these events' `__isBroken` internal field to `false`.

Event Breaker Types

Several types of Event Breaker can be applied to incoming data streams:

- Regex

- File Header
- JSON Array
- JSON New Line Delimited
- Timestamp
- CSV

Regex

The Regex breaker uses regular expressions to find breaking points in data streams.

After a breaker regex pattern has been selected, it will apply on the stream **continuously**. Breaking will occur at the beginning of the match, and the matched content will be consumed/thrown away. If necessary, you can use a positive lookahead regex to keep the content – e.g.: `(?=pattern)`

Capturing groups are **not allowed** to be used anywhere in the Event Breaker pattern, as they will further break the stream – which is often undesirable. Breaking will also occur if the final broken event's length reaches the rule's **Max event bytes**.

Example

Break after a newline or carriage return, but only if followed by a timestamp pattern:
Event Breaker: `[\n\r]+(?=\d+-\d+-\d+\s\d+:\d+:\d+)`

Sample Event - Multiline

```
--- input ---
2020-05-19 16:32:12 moen3628 ipsum[5213]: Use the mobile TCP feed, then you can pr
   Try to connect the FTP sensor, maybe it will connect the digital bus!
   Try to navigate the AGP panel, maybe it will quantify the mobile alarm!
2020-05-19 16:32:12 moen3628 ipsum[5213]: Use the mobile TCP feed, then you can pr
   Try to connect the FTP sensor, maybe it will connect the digital bus!
   Try to navigate the AGP panel, maybe it will quantify the mobile alarm!


--- output event 1 ---
{
  "_raw": "2020-05-19 16:32:12 moen3628 ipsum[5213]: Use the mobile TCP feed, then
  "_time": 1589920332
}

--- output event 2 ---
{
  "_raw": "2020-05-19 16:32:12 moen3628 ipsum[5213]: Use the mobile TCP feed, then
  "_time": 1589920332
}
```

File Header

You can use the File Header breaker to break files with headers, such as IIS or Bro logs. This type of breaker relies on a header section that lists field names. The header section is typically present at the top of the file, and can be single-line or greater.

After the file has been broken into events, fields will also be extracted, as follows:

- **Header Line**: Regex matching a file header line. For example, `^#`.
- **Field Delimiter**: Field delimiter regex. For example, `\s+`.
- **Field Regex**: Regex with one capturing group, capturing all the fields to be broken by field delimiter. For example, `^#[Ff]ields[:]?\s+(.*)`
- **Null Values**: Representation of a null value. Null fields are not added to events.
- **Clean Fields**: Whether to clean up field names by replacing non `[a-zA-Z0-9]` characters with `_`.

Example

Using the values above, let's see how this sample file breaks up:

Sample Event - File Header

```
--- input ---
#fields ts      uid     id.orig_h       id.orig_p       id.resp_h       id.resp_p
#types  time    string  addr    port    addr    port    enum
1331904608.080000       -     192.168.204.59  137     192.168.204.255 137     udp
1331904609.190000       -     192.168.202.83  48516   192.168.207.4   53      udp


--- output event 1 ---
{
  "_raw": "1331904608.080000       -     192.168.204.59  137     192.168.204.255 13
  "ts": "1331904608.080000",
  "id_orig_h": "192.168.204.59",
  "id_orig_p": "137",
  "id_resp_h": "192.168.204.255",
  "id_resp_p": "137",
  "proto": "udp",
  "_time": 1331904608.08
}

--- output event 2 ---
{
  "_raw": "1331904609.190000       -     192.168.202.83  48516   192.168.207.4   53
  "ts": "1331904609.190000",
  "id_orig_h": "192.168.202.83",
  "id_orig_p": "48516",
  "id_resp_h": "192.168.207.4",
  "id_resp_p": "53",
  "proto": "udp",
  "_time": 1331904609.19
}
```

JSON Array

You can use the JSON Array to extract events from an array in a JSON document (e.g., an Amazon CloudTrail file).

- **Array Field**: Optional path to array in a JSON event with records to extract. For example, `Records`.
- **Timestamp Field**: Optional path to timestamp field in extracted events. For example, `eventTime` or `level1.level2.eventTime`.

- **JSON Extract Fields**: Enable this toggle to auto-extract fields from JSON events. If disabled, only `_raw` and `time` will be defined on extracted events.
- **Timestamp Format**: If **JSON Extract Fields** is set to `No`, you **must** set this to **Autotimestamp** or **Current Time**. If **JSON Extract Fields** is set to `Yes`, you can select any option here.

Example

Using the values above, let's see how this sample file breaks up:

Sample Event - JSON Document (Array)

```
--- input ---
{"Records":[{"eventVersion":"1.05","eventTime":"2020-04-08T01:35:55Z","eventSource"
{"eventVersion":"1.05","eventTime":"2020-04-08T01:35:56Z","eventSource":"ec2.amazo

--- output event 1 ---
{
  "_raw": "{\"eventVersion\":\"1.05\",\"eventTime\":\"2020-04-08T01:35:55Z\",\"ever
  "_time": 1586309755,
  "cribl_breaker": "j-array"
}

--- output event 2 ---
{
  "_raw": "{\"eventVersion\":\"1.05\",\"eventTime\":\"2020-04-08T01:35:56Z\",\"ever
  "_time": 1586309756,
  "cribl_breaker": "j-array"
}
```

JSON New Line Delimited

You can use the JSON New Line Delimited breaker to break and extract fields in newline-delimited JSON streams.

Example

Using default values, let's see how this sample stream breaks up:

Sample Event - Newline Delimited JSON Breaker

```
--- input ---
{"time":"2020-05-25T18:00:54.201Z","cid":"w1","channel":"clustercomm","level":"info
{"time":"2020-05-25T18:00:54.246Z","cid":"w0","channel":"clustercomm","level":"info


--- output event 1 ---
{
  "_raw": "{\"time\":\"2020-05-25T18:00:54.201Z\",\"cid\":\"w1\",\"channel\":\"clus
  "time": "2020-05-25T18:00:54.201Z",
  "cid": "w1",
  "channel": "clustercomm",
  "level": "info",
  "message": "metric sender",
  "total": 720,
  "dropped": 0,
  "_time": 1590429654.201,
}

--- output event 2 ---
{
  "_raw": "{\"time\":\"2020-05-25T18:00:54.246Z\",\"cid\":\"w0\",\"channel\":\"clus
  "time": "2020-05-25T18:00:54.246Z",
  "cid": "w0",
  "channel": "clustercomm",
  "level": "info",
  "message": "metric sender",
  "total": 720,
  "dropped": 0,
  "_time": 1590429654.246,
}
```

Timestamp

You can use the Timestamp breaker to break events at the beginning of any line in which Cribl Edge finds a timestamp. This type enables breaking on lines whose timestamp pattern is not known ahead of time.

Example

Using default values, let's see how this sample stream breaks up:

Sample Event - Timestamp Based Breaker

```
--- input ---
{"level":"debug","ts":"2021-02-02T10:38:46.365Z","caller":"sdk/sync.go:42","msg":"H
{"level":"debug","ts":"2021-02-02T10:38:56.365Z","caller":"sdk/sync.go:42","msg":"H


--- output event 1 ---
{
  "_raw": "{\"level\":\"debug\",\"ts\":\"2021-02-02T10:38:46.365Z\",\"caller\":\"sd
  "_time": 1612262326.365
}

--- output event 2 ---
{
  "_raw": "{\"level\":\"debug\",\"ts\":\"2021-02-02T10:38:56.365Z\",\"caller\":\"sd
  "_time": 1612262336.365
}
```

CSV

The CSV breaker extracts fields in CSV streams that include a header line. Selecting this type exposes these extra fields:

- **Delimiter**: Delimiter character to use to split values. Defaults to: `,`
- **Quote Char**: Character used to quote literal values. Defaults to: `"`
- **Escape Char**: Character used to escape the quote character in field values. Defaults to: `"`
  Example: Using default values, let's see how this sample stream breaks up:

Example

Using default values, let's see how this sample stream breaks up:

Sample Event - CSV Breaker

```
--- input ---
time,host,source,model,serial,bytes_in,bytes_out,cpu
1611768713,"myHost1","anet","cisco","ASN4204269",11430,43322,0.78
1611768714,"myHost2","anet","cisco","ASN420423",345062,143433,0.28


--- output event 1 ---
{
  "_raw": "\"1611768713\",\"myHost1\",\"anet\",\"cisco\",\"ASN4204269\",\"11430\",\
  "time": "1611768713",
  "host": "myHost1",
  "source": "anet",
  "model": "cisco",
  "serial": "ASN4204269",
  "bytes_in": "11430",
  "bytes_out": "43322",
  "cpu": "0.78",
  "_time": 1611768713
}

--- output event 2 ---
{
  "_raw": "\"1611768714\",\"myHost2\",\"anet\",\"cisco\",\"ASN420423\",\"345062\",\
  "time": "1611768714",
  "host": "myHost2",
  "source": "anet",
  "model": "cisco",
  "serial": "ASN420423",
  "bytes_in": "345062",
  "bytes_out": "143433",
  "cpu": "0.28",
  "_time": 1611768714
}
```

With **Type: CSV** selected, an Event Breaker will properly add quotes around all values, regardless of their initial state.

Cribl Versus Custom Rulesets

Event Breaker rulesets shipped by Cribl will be listed under the **Cribl** tag, while user-built rulesets will be listed under **Custom**. Over time, Cribl will ship more patterns, so this distinction allows for both sets to grow independently. In the case of an ID/Name conflict, the Custom pattern takes priority in listings and search.

Exporting and Importing Rulesets

You can export and import Custom (or Cribl) rulesets as JSON files. This facilitates sharing rulesets among Worker Groups or Cribl Edge deployments.

To export a ruleset:

1. Click to open an existing ruleset, or create a new ruleset.
2. In the resulting modal, click **Advanced Mode** to open the JSON editor.
3. You can now modify the ruleset directly in JSON, if you choose.
4. Click **Export**, select a destination path, and name the file.

To import any ruleset that has been exported as a valid JSON file:

1. Create a new ruleset.
2. In the resulting modal, click **Advanced Mode** to open the JSON editor.
3. Click **Import**, and choose the file you want.
4. Click **OK** to get back to the **New Ruleset** modal.
5. Click **Save**.

Every ruleset must have a unique value in its top `id` key. Importing a JSON file with a duplicate `id` value will fail at the final **Save** step, with a message that the ruleset already exists. You can remedy this by giving the file a unique `id` value.

Edge: Building Breakers from Files

You can use Cribl Edge to apply and author Event Breakers while exploring files – even files that you have not saved as sample files. This option includes remote files that are viewable only from Edge. See Exploring Cribl Edge on Linux or Exploring Cribl Edge on Windows.

Troubleshooting

If you notice fragmented events, check whether Cribl Edge has added a `__timeoutFlush` internal field to them. This diagnostic field's presence indicates that the events were flushed because the Event Breaker

buffer timed out while processing them. These timeouts can be due to large incoming events, backpressure, or other causes.

Specifying Minimum `_raw` Length

If you notice that the same Collector or Source is applying inconsistent Event Breakers to different datasets – e.g., intermittently falling back to the **System Default Rule** – you can address this by adjusting Breakers' **Min Raw Length** setting.

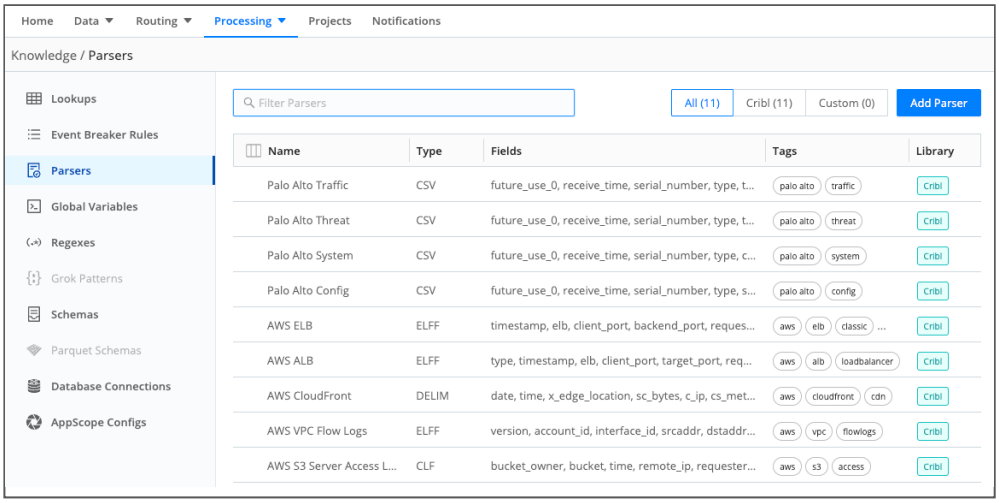
When determining which Event Breaker to use, Cribl Edge creates a substring from the incoming `_raw` field, at the length set by rules' **Max Event Bytes** parameters. If Cribl Edge does not get a match, it checks whether the substring of `_raw` has at least the threshold number of characters set by **Min Raw Length**. If so, Cribl Edge does not wait for more data.

You can tune this **Min Raw Length** threshold to account for varying numbers of inapplicable characters at the beginning of events. The default is `256` characters, but you can set this as low as `50`, or as high as `100000` (100 KB).

15.3. Parsers Library

What Are Parsers

Parsers are definitions and configurations for the Parser Function. You can find the library from Cribl Edge's top nav under **Processing** > **Knowledge** > **Parsers**, and its purpose is to provide an interface for creating and editing Parsers. The library is searchable, and each parser can be tagged as necessary.



Parsers Library

Parsers can be used to **extract** or **reserialize** events. See Parser Function page for examples.

Supported Parser Types:

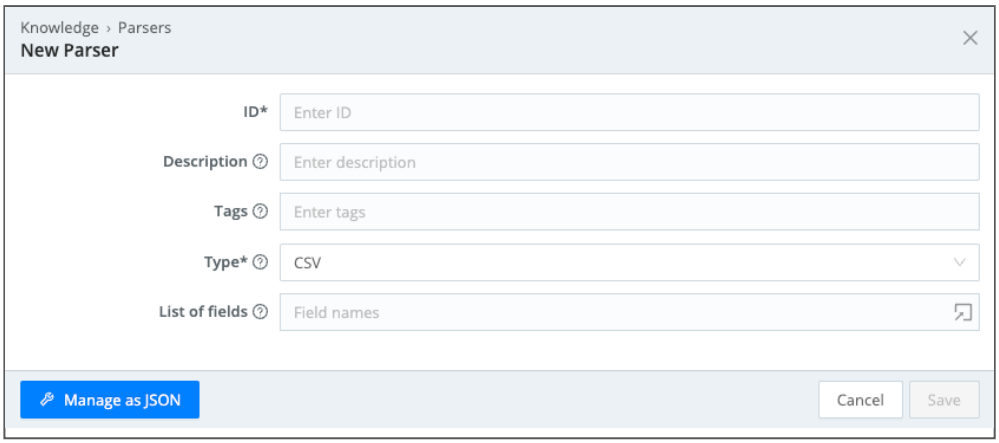
- CSV – Parse and reserialize comma-separated values.
- ELFF – Parse and reserialize events in Extended Log File Format.
- CLF – Parse and reserialize events in Common Log Format.

Creating a Parser

To create a parser, follow these steps:

1. Go to **Knowledge** > **Parsers** and click **Add Parser**.
2. Enter a unique **ID**.
3. Optionally, enter a **Description**.
4. Select a **Type** (see the supported types above).

5. Enter the **List of fields** expected to be extracted, in order. Click this field's Maximize icon (far right) if you'd like to open a modal where you can work with sample data and iterate on results.
6. Optionally, enter any desired **Tags**.



Adding a new parser

15.4. Global Variables Library

What Are Global Variables

Global Variables are reusable JavaScript expressions that can be accessed in Functions in any Pipeline. You can access the library from Cribl Edge's top nav under **Processing** > **Knowledge** > **Global Variables**.

Typical use cases for Global Variables include:

- Storing a constant that you can reference from any Function in any Pipeline.
- Storing a relatively long value expression, or one that uses one or more **arguments**.

Global Variables can be of the following types:

- Number
- String
- Boolean
- Object
- Array
- Expression

Global Variables can be accessed via `C.vars.` – which can be called anywhere in Cribl Edge that JS expressions are supported. Typeahead is provided. More on Cribl Expressions here.

Examples

Scenario 1:

Assign field `foo` the value in `theAnswer` Global Variable.

- Global Variable Name: `theAnswer` <– ships with Cribl Edge by default.
- Global Variable Value: `42`
- **Sample Eval Function:** `foo = C.vars.theAnswer`

Scenario 2:

Assign field `nowEpoch` the current time, in epoch format.

- Global Variable Name: `epoch` <– ships with Cribl Edge by default.
- Global Variable Value: `Date.now()/1000`
- **Sample Eval Function:** `nowEpoch = C.vars.epoch()`

Scenario 3:

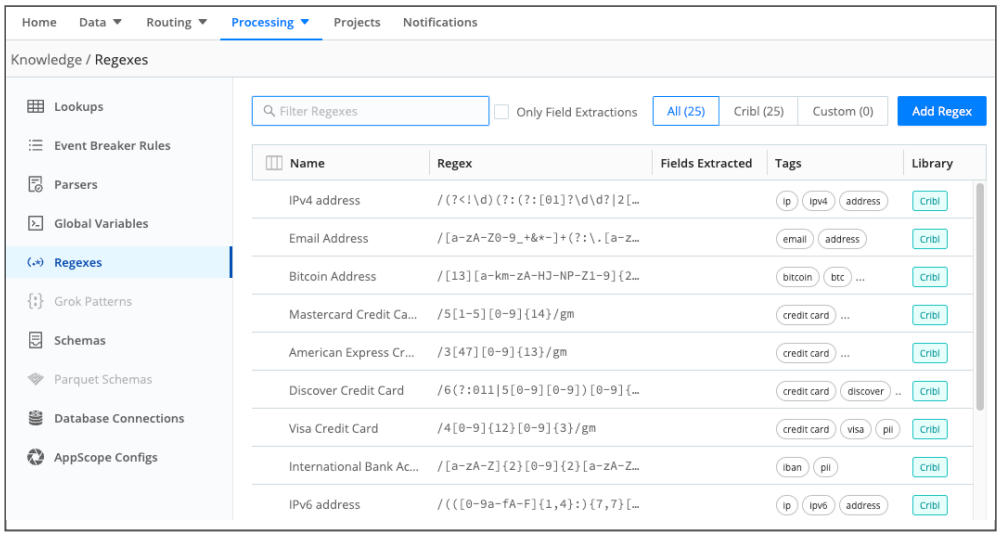
Create a new field called `storage`, by converting the value of event field `size` to human-readable format.

- Global Variable Name: `convertBytes` <– ships with Cribl Edge by default
- Global Variable Value: `` `${Math.round(bytes / Math.pow(1024, (Math.floor(Math.log(bytes) / Math.log(1024)))), 2)}${['Bytes', 'KiB', 'MiB', 'GiB', 'TiB', 'PiB', 'EiB', 'ZiB', 'YiB'][(Math.floor(Math.log(bytes) / Math.log(1024)))]}` ``
  Note the use of quotes or backticks around values. Use the opposite delimiter for the enclosing expression.
- Global Variable Argument: `bytes`
- **Sample Eval Function:** `storage = C.vars.convertBytes(size)`
  Note the use of `bytes` here as an argument.

15.5. Regex Library

What Is the Regex Library

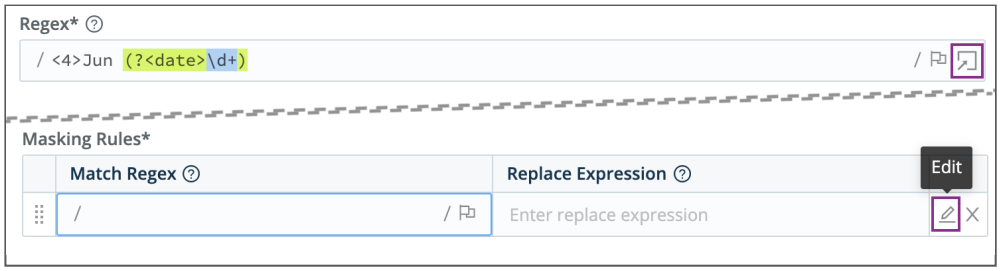
Cribl Edge ships with a Regex Library that contains a set of pre-built common regex patterns. This library serves as an easily accessible repository of regular expressions. The Library is searchable, and you can assign tags to each pattern for further organization or categorization. Access the Library from Cribl Edge's top nav under **Processing** > **Knowledge** > **Regexes** .



Regular Expression Library

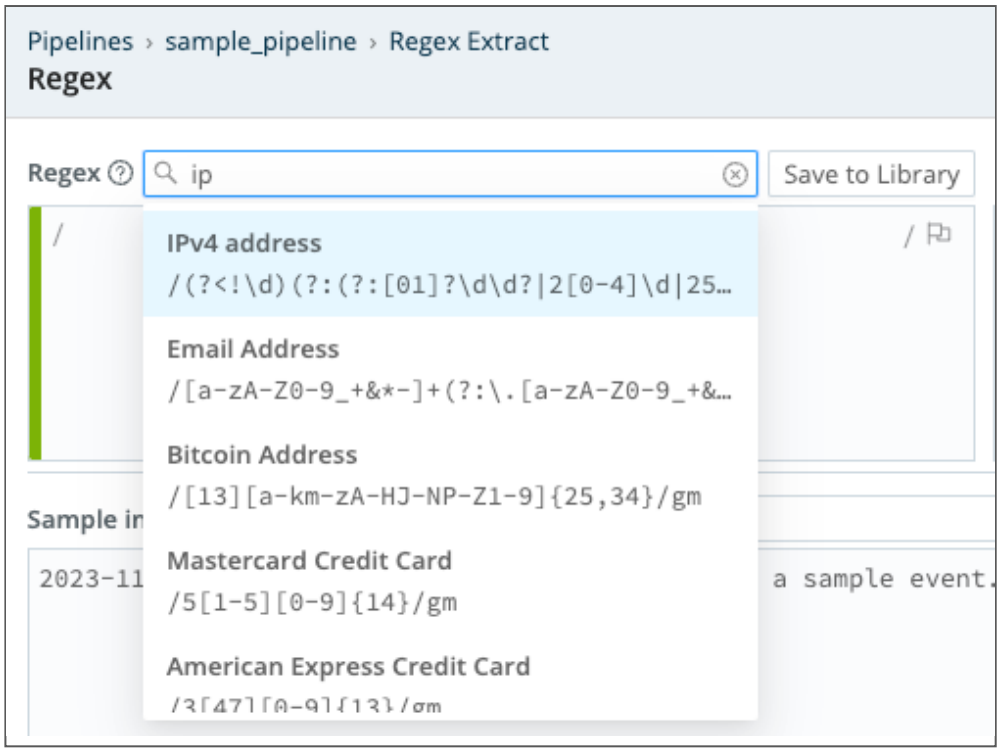
Using Library Patterns

As of this version, the Library contains 25 patterns shipped by Cribl Edge. To insert a pattern into a Function's regex field, first click the pop-out or Edit icon beside that field.



Opening a Regex modal

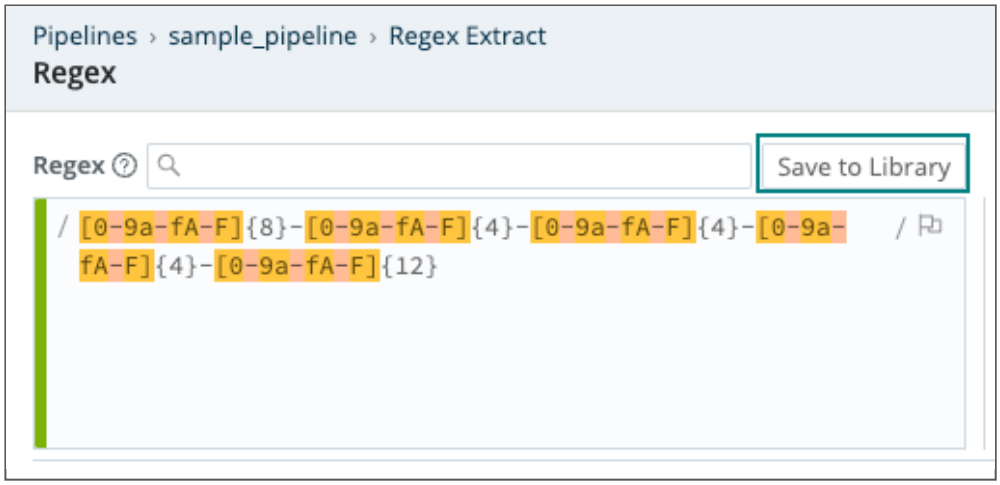
In the resulting Regex or Rules modal, Regex Library patterns will appear as typeahead options. Click a pattern to paste it in. You can then use the pattern as-is, or modify it as necessary.



Inserting a pattern from the Regex Library

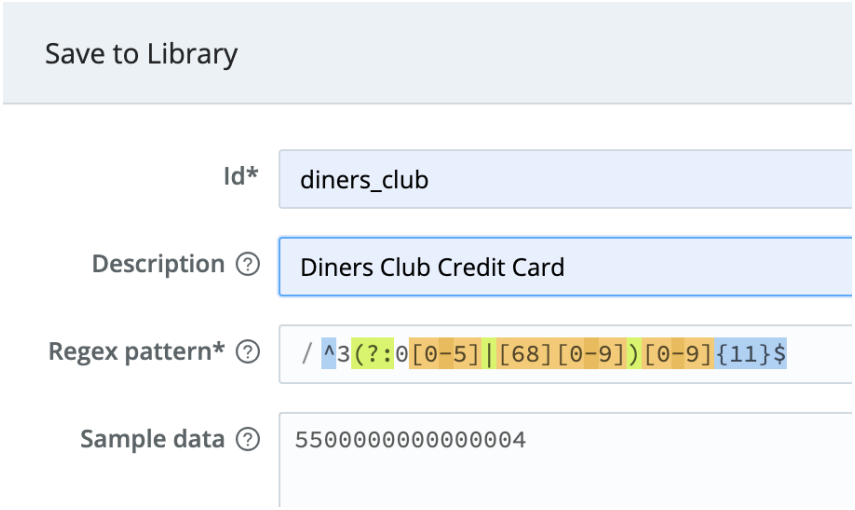
Adding Patterns to the Library

You can also add new, custom patterns to the Library. In the same modal, once you've built your pattern, click the **Save to Library** button.



Adding a custom pattern to the Regex Library from a Function's Regex modal

In the resulting modal, give your custom pattern a unique ID. Optionally, you can also provide a **Description** (name) and groom the **Sample data**. Then click **Save**.



Identifying the custom pattern

Your custom pattern will now reside in the Regex Library. It will be available to Functions using the same typeahead assist as Cribl's pre-built patterns.

Cribl vs. Custom and Priority

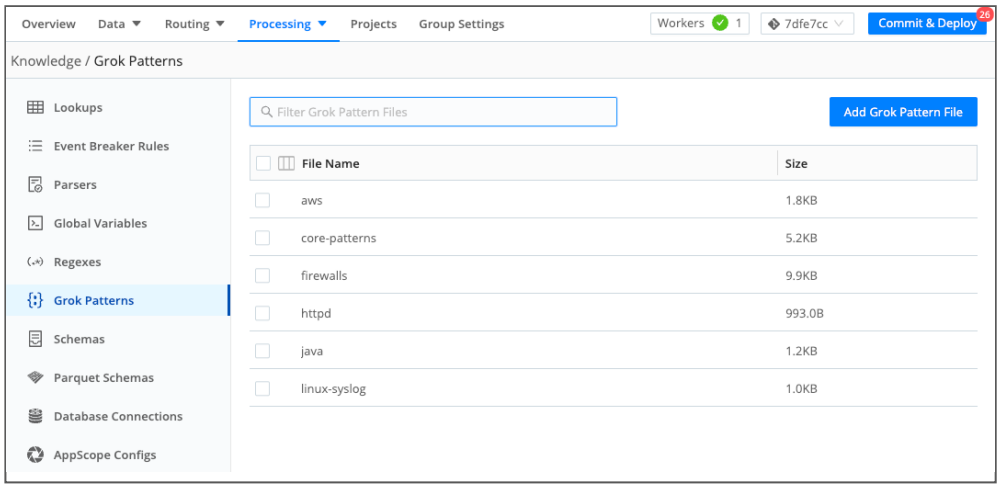
Within the Library, patterns shipped by Cribl will be listed under the **Cribl** tab, while those built by users will be found under **Custom**. Over time, Cribl Edge will ship more patterns, and this distinction allows for both sets to grow independently.

In the case of an ID/Name conflict, the Custom pattern takes priority in listings and search. For example, if a Cribl-provided pattern and a Custom one are both named `ipv4`, the one from Cribl will not be displayed or delivered as a search result.

15.6. Grok Patterns Library

What Is the Grok Patterns Library

Cribl Edge ships with a Grok Patterns Library that contains a set of pre-built common patterns, organized as files.



Grok Patterns Library

Managing Library Patterns

You can access the Grok Patterns Library from the UI's top nav by selecting **Processing** > **Knowledge** > **Grok Patterns**. The library contains several pattern files that Cribl provides for basic Grok scenarios, and is searchable.

To edit a pattern file, click **Edit** in its **Actions** column.

To create a new pattern file, click **Add Grok Pattern File**. In the resulting modal, assign a unique **Filename**, populate the file with patterns, then click **Save**.

Pattern files reside in: `$CRIBL_HOME/(default|local)/cribl/grok-patterns/`

Using Grok Patterns

In the current Cribl Edge version, you apply Grok patterns by inserting a Grok Function into a Pipeline, then manually typing or pasting patterns into the **Pattern** field(s).

15.7. Schemas Library

Cribl Edge supports two kinds of schemas:

- JSON schemas for validating JSON events. This page outlines how to manage these in the UI at **Knowledge** > **Schemas**.
- Parquet schemas for writing data from a Cribl Edge Destination to Parquet files.

These schemas obviously serve different purposes. Beware of assuming that operations that work for one kind of schema can be used with the other. For example, the validation method for JSON schemas – `C.Schema('<schema_name>').validate(<object_field>)` – can't be used to validate Parquet schemas.

JSON Schemas

JSON schemas are based on the popular JSON Schema standard, and Cribl Edge supports schemas matching that standard's Drafts 0 through 7.

You can access the schema library from Cribl Edge's top nav under **Processing** > **Knowledge** > **Schemas**. Its purpose is to provide an interface for creating, editing, and maintaining schemas.

You validate a schema using this built-in method:
`C.Schema('<schema_name>').validate(<object_field>)`.

You can call this method anywhere in Cribl Edge that supports JavaScript expressions. Typical use cases for schema validation:

- Making a decision before sending an event down to a destination.
- Making a decision before accepting an event. (E.g., drop an event if invalid.)
- Making a decision to route an event based on the result of validation.

JSON Schema Example

To add this example JSON Schema, go to **Processing** > **Knowledge** > **Schemas** and click **New Schema**. Enter the following:

- ID: `schema1`.
- Description: (Enter your own description here.)
- Schema: Paste the following schema.

JSON Schema - Sample

```
{
  "$id": "https://example.com/person.schema.json",
  "$schema": "http://json-schema.org/draft-07/schema#",
  "title": "Person",
  "type": "object",
  "required": ["firstName", "lastName", "age"],
  "properties": {
    "firstName": {
      "type": "string",
      "description": "The person's first name."
    },
    "lastName": {
      "type": "string",
      "description": "The person's last name."
    },
    "age": {
      "description": "Age in years which must be equal to or greater than zero.",
      "type": "integer",
      "minimum": 0,
      "maximum": 42
    }
  }
}
```

Assume that events look like this:

Events

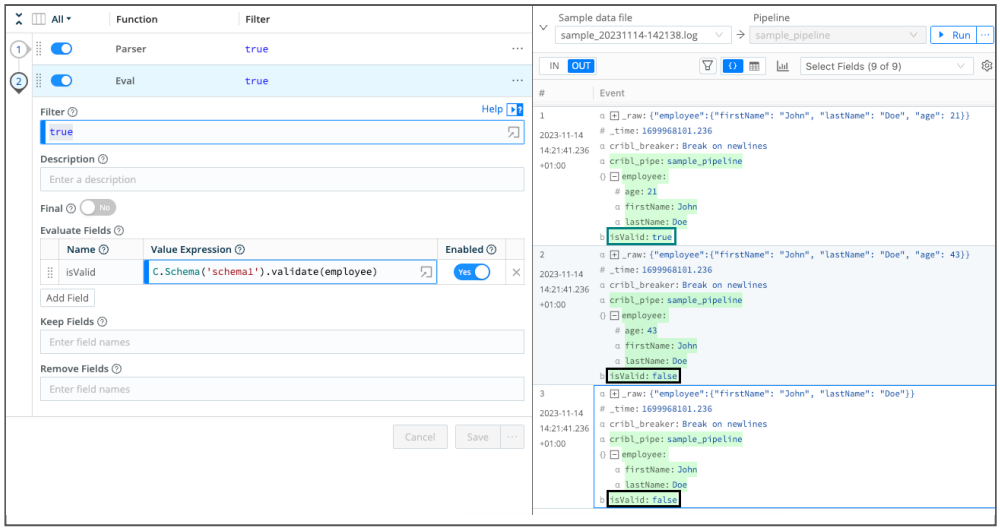
```
{"employee":{"firstName": "John", "lastName": "Doe", "age": 21}}
{"employee":{"firstName": "John", "lastName": "Doe", "age": 43}}
{"employee":{"firstName": "John", "lastName": "Doe"}}
```

To validate whether the `employee` field is valid per `schema1`, we can use the following:
`C.Schema('schema1').validate(employee)`

Results:

- First event is **valid**.
- Second event is **not valid** because `age` is greater than the maximum of `42`.

- Third event is **not valid** because `age` is missing.



Schema validation results for the above events

15.8. Parquet Schemas

Cribl Edge supports two kinds of schemas:

- Parquet schemas for writing data from a Cribl Edge Destination to Parquet files. This page outlines how to manage these in the UI at **Knowledge** > **Parquet Schemas**.
- JSON schemas for validating JSON events.

These schemas obviously serve different purposes. Beware of assuming that operations that work for one kind of schema can be used with the other. For example, the validation method for JSON schemas – `C.Schema('<schema_name>').validate(<object_field>)` – can't be used to validate Parquet schemas.

Configuring Parquet Schemas

Destinations whose **General Settings** > **Data format** drop-down includes a `Parquet` option can write out data as files in the Apache Parquet columnar storage format.

When logging is set to `debug`, you will see what schema is associated with the Parquet files that Cribl Edge writes out. On Linux, you can use the Cribl Edge CLI's `parquet` command to view a Parquet file, its metadata, or its schema.

If you turn on a Destination's **Parquet Settings** > **Automatic schema**, Cribl Edge will automatically generate a Parquet schema based on the events of each Parquet file it writes. The automatically-generated schema preserves the exact structure of the ingested events, with no data lost. In Cribl Edge v.4.4.2, every field in the generated schema will have a data type of `JSON`, `STRING`, `INT_64`, `UNSIGNED_INT_64`, or `DOUBLE`.

(The one exception is the Amazon Security Lake Destination, where automatic schema generation is not relevant because Amazon Security Lake exclusively uses the OCSF Parquet schemas already available in the drop-down.)

Pros and Cons of Automatically Generated Parquet schemas

When you decide whether or not to generate Parquet schemas automatically, consider these pros and cons.

Advantages:

- Configuration is minimal.
- There's no need to know how events you want to ingest are structured.
- There's no need to change Parquet schema when event structure changes.

Disadvantages:

- Writing Parquet files takes longer than with predefined schemas. Note that schema complexity increases linearly with the number of events used to generate the schema.
- Automatically-generated Parquet schemas cannot be as fine-tuned as predefined ones, to capture certain nuances. For example, encoding will always be `PLAIN`, compression will always be `SNAPPY`.
- Some aspects of event structure (such as maps and sets) cannot be captured in the schema.
- Some aspects of time-related fields cannot be captured in the schema.
- Because the system produces a new Parquet schema for each file written, there may be differences between successive schemas. This can limit your ability to concatenate a collection of Parquet files.

Configuring Parquet Schemas

If you are not using automatic schema generation, or if there's no pre-existing schema that suits your data:

- Before configuring a Destination for Parquet output, you should add an existing Parquet schema, or create a new Parquet schema that suits the data you're working with.
- You do not need to start from scratch: Cribl provides sample Parquet schemas for you to clone and then customize as needed.

From the top nav, select **Processing** > **Knowledge**, then select the **Parquet Schemas** left tab.

If you're adding a Parquet schema:

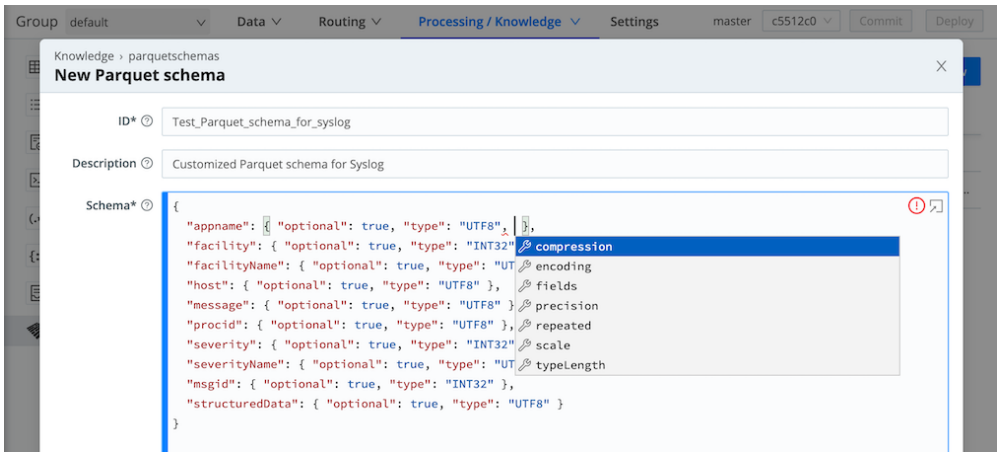
- Click **Add Parquet schema** to open the **New Parquet schema** modal.

If you're cloning an existing Parquet schema to create a new schema:

1. Click the name of the schema you want to start with, to open it in a modal.
2. Click **Clone Parquet Schema** to open the **New Parquet schema** modal.
3. Give the new schema a name and description.

From here, whether you're working with a brand-new or cloned schema:

4. Add and/or edit fields as desired.
5. Click **Save**.



Creating a Parquet schema

Now, when you configure your Destination, the schema you created will be available from the **Parquet Settings** > **Parquet schema** drop-down.

Creating/Modifying Parquet Schemas from Destinations

When the existing schemas on a Destination's **Parquet schema** drop-down don't meet your needs, follow this procedure to add or modify (clone) a new schema.

1. Navigate to the Parquet Schemaa Library at **Processing** > **Knowledge** > **Parquet Schemas**.
2. Add or modify the schema, as described above.

To modify an existing schema, Cribl strongly recommends that you first clone the schema, and give the clone its own distinct name. See Cloning Is Safer just below.

3. Return to the Destination, and select the new or modified schema from the **Parquet schema** drop-down.

Cloning Is Safer than Modifying

Modifying an existing schema in the Schema Library does **not** propagate your modifications to the Destination's copy of the schema.

- Cloning and renaming schemas is the safest approach, because in Step 3 above, this ensures that your Destination will now use the newly modified version of the schema.
- If you do **not** clone and rename the schema (i.e., you leave the schema name unchanged), you still must **re-select** the schema in the Destination's **Parquet schema** drop-down to bring the modified version into the Destination.

Working with Parquet in Cribl Edge

Different Parquet readers and writers behave differently. Keep the following guidelines in mind when working with Parquet in Cribl Edge.

The Parquet Schema Editor

In the Parquet Schema editor, you express your Parquet schema in JSON. This does not look like the examples in the Parquet spec, which have their own syntax (like the repetition/name/type triple), but they are functionally equivalent. See the examples below.

The editor provides autocompletion and validation to guide you. (However, Cribl Edge currently does not fully support autocompletion on deeply nested schemas.)

File Extensions

For now, Cribl Edge can read Parquet files only if they have the extension `.parquet`, `.parq`, or `.pqt`.

Field Content

When Cribl Edge writes to a Parquet file:

- If the data contains a field that **is not** present in the schema – i.e, an extra field – Cribl Edge writes out the parent rows, but omits the extra field.
- If the data contains a field that **is** present in the Parquet schema, but whose properties violate the schema, Cribl Edge treats this as a data mismatch. Cribl Edge drops the rows containing that field – it does not write those rows to the output Parquet file at all.
- If the data contains JSON, the JSON must be stringified. Otherwise, Cribl Edge treats this as a data mismatch, and does not write out the row. For example, this valid (but not stringified) JSON will trigger a data mismatch:
  `{ "name": "test"}`.
  The same JSON in stringified form will work fine:
  `"{\"name\": \"test\"}"`.

Data Types

Cribl Edge supports:

- All primitive types.
- All logical types.
- All converted types.

Converted types have been superseded by logical types, as described in the Apache Parquet docs. Cribl Edge can read Parquet files that use converted types, but will write out the same data using corresponding logical types.

Repetition Type

You have three alternatives when defining a field's Repetition type:

- Set `optional` to `true`.
- Set `repeated` to `true`.
- Set neither `optional` nor `repeated`. This **implicitly** sets the Repetition type to `required`, and it is the default.

Usage guidelines:

- Do not set both `optional` **and** `repeated` to `true`.
- Do not use the `required` key at all.
- Instead of omitting `optional`, you have the option to include it, but set it to `false`.
- Instead of omitting `repeated`, you have the option to include it, but set it to `false`.
- If any field's Repetition type is `repeated`, Cribl Edge represents this field as a single key whose value is an array – not as separate key-value pairs with identical keys.

Encodings

- Among the `*DICTIONARY` encodings, Cribl Edge supports only `DICTIONARY`. Trying to assign the unsupported encodings `PLAIN_DICTIONARY` or `RLE_DICTIONARY` will produce an error.
- `BYTE_STREAM_SPLIT` encoding can be used only with `DOUBLE` or `FLOAT` types, and otherwise produces errors.
- The `RLE` and all `DELTA*` encodings also produce errors.

Parquet Schema Expressed as JSON

These examples should give you a hint of how to express Parquet schema in JSON. For a full explanation of the Parquet syntax, see the Parquet spec.

List Example

In this example, the whole schema is named `the_list`, and its structure consists of a container named `list` whose zero or more elements are each named `element`. Note that in Parquet schema, to specify that

a field is nested (i.e., contains other fields), you just give it the type `group`.

Parquet Schema Example 1

```
message schema {
  REQUIRED group the_list (LIST) {
    REPEATED group list {
      REQUIRED BYTE_ARRAY element (STRING);
    }
  }
}
```

Expressed as JSON, the same schema is more spread-out, because to specify that a field is nested, you give it a sub-field named `fields`. (This is equivalent to declaring the nested field's type as `group`.)

Parquet Schema Example 1 Expressed as JSON

```
{
  "the_list": {
    "type": "LIST",
    "fields": {
      "list": {
        "repeated": true,
        "fields": {
          "element": {
            "type": "STRING"
          }
        }
      }
    }
  }
}
```

Map Example

In this example, the whole schema is named `the_map`, and its structure consists of a container named `key_value` whose zero or more pairs of elements each contain a `key` and a `value` element. Both the `key` and the `value` are strings in this case, but they could be other data types.

Parquet Schema Example 2

```
message schema {
  REQUIRED group the_map (MAP) {
    REPEATED group key_value {
      REQUIRED BYTE_ARRAY key (STRING);
      REQUIRED BYTE_ARRAY value (STRING);
    }
  }
}
```

Parquet Schema Example 2 Expressed as JSON

```
{
  "the_map": {
    "type": "MAP",
    "fields": {
      "key_value": {
        "repeated": true,
        "fields": {
          "key": {
            "type": "STRING"
          },
          "value": {
            "type": "STRING"
          }
        }
      }
    }
  }
}
```

Array of Arrays Example

In this example, the whole schema is named `array_of_arrays`, and its structure consists of a container named `list` whose zero or more elements – which are themselves lists – are each named `element`. Within each of those `element` lists are one or more additional `list/element` structures.

Parquet Schema Example 3

```
message schema {
  REQUIRED group array_of_arrays (LIST) {
    REPEATED group list {
      REQUIRED group element (LIST) {
        REPEATED group list {
          REQUIRED INT64 element;
        }
      }
    }
  }
}
```

Expressed as JSON, the same schema is cumbersome to read compared to Parquet, which was designed to express nested structures concisely.

Parquet Schema Example 3 Expressed as JSON

```
{
  "array_of_arrays": {
    "type": "LIST",
    "fields": {
      "list": {
        "repeated": true,
        "fields": {
          "element": {
            "type": "LIST",
            "fields": {
              "list": {
                "repeated": true,
                "fields": {
                  "element": {
                    "type": "INT64"
                  }
                }
              }
            }
          }
        }
      }
    }
  }
}
```

Limitations

Cribl Edge does not support writing Parquet files which employ any of the following:

- Parquet Modular Encryption.
- The Parquet Bloom Filter.
- Separation of metadata and/or column data into multiple files.

Cribl Edge does not support reading or writing Parquet files which employ the deprecated `INT96` data type.

15.9. AppScope Configs

You can obtain high-fidelity application data from any Linux command or application using AppScope, an open-source tool from Cribl. This includes data that's difficult or impossible to obtain any other way. (Note that AppScope is no longer being actively developed by Cribl.)

AppScope is included in your Cribl Edge installation. One or two AppScope Sources are enabled right out-of-the-box; they are configured to send data to Cribl Edge or Cribl Stream on the same host.

AppScope ships with a set of config files that offer a variety of options for what events and metrics you want AppScope to capture, and whether to enable payload capture. These are: `sample_config`, `default_metrics_events`, `verbose_metrics_events`, and `verbose_metrics_events_payloads`.

To understand precisely what one of these files does, select it to open its **AppScope Configs** modal and click **Manage as JSON**. For explanations and definitions, see the Config File and Schema Reference sections in the AppScope docs.

You can customize the config files, or edit them via **Manage as JSON**. When editing, you can revert to the original state by clicking **Restore Default**. The safer option is to click **Clone AppScope config**, then edit and save your new version with a new filename. The sample configs, and any variants you have created and saved, constitute a reusable set of AppScope configurations in the AppScope Config Library.

When you have assigned an AppScope config to a Rule in an AppScope Source, and you change it in the Knowledge UI, the changes will not take effect until you do the following:

- In the AppScope Source **AppScope Rules** settings, re-assign the config to the Rule. (That is, unselect the config file, then select it again.)
- Save and commit.

Creating AppScope Configs

You access the **AppScope Configs** UI from Cribl Edge's top nav:

- In Cribl Stream, select **Processing** > **Knowledge** > **AppScope Configs**.
- In Cribl Edge, select **More** > **Knowledge** > **AppScope Configs**.

Click **Add AppScope config** to open a modal to begin creating an AppScope config, using the options below.

Once you've created AppScope configs, you can edit, delete, search, and tag them as desired.

General Settings

ID: Enter a unique name to identify this AppScope Config.

Description: Optionally, enter a description that will remind you what the config is for when you see it listed in the library.

Tags: Optionally, enter any desired tags.

Transports

AppScope offers several options for routing data. See Data Routing in the AppScope docs.

Use AppScope Source's IP/port or UNIX domain socket?: The default `Yes` setting applies the transport options defined in the associated AppScope Source. Toggling this to `No` exposes the following fields in this section.

Cribl authentication token: Optionally, set a value to add to the `config-event` header (as a top-level `authToken` property) when AppScope establishes a connection.

Send metrics and events over the same connection?: Defaults to `Yes`, which locks some settings into preconfigured values. Toggle to `No` if you prefer to route metrics and events independently.

Connection type: From the drop-down, choose one of the connection types described below, then enter values in whatever fields the UI exposes.

- When **Send metrics and events over the same connection?** is toggled to `Yes`, you configure a single connection by choosing `TCP`, `Edge` (the default), or `Unix`.
- When routing metrics and events independently, you configure one connection for metrics and another for events; the drop-down offers two additional connection types: `UDP` and `File`.

Transports and Connection Types

At the heart of any AppScope config are the transports you define. In AppScope, a transport specifies something to send, how to send it, and where to send it. The things to send can be metrics, events, or logs.

Depending on the choices you make in the **Transport** settings, the UI will make some or all of the following types of connections available.

UDP: The destination is the network server whose hostname or IP address you specify in the **Host** field, along with a UDP **Port** to listen on.

TCP: The destination is the network server whose hostname or IP address you specify in the **Host** field, along with a TCP **Port** to listen on. The **Enable TLS** toggle defaults to `No`.

Unix: The destination is a Unix domain server listening on the Unix domain **Socket path** you specify.

File: The destination is the file whose directory location you specify in the **File path** field, along with a **Buffering** method.

Edge: The destination is Cribl Edge, listening on a preconfigured Unix domain socket.

TLS Settings (TCP Only)

When the **Connection type** is `TCP`, the UI exposes the **Enabled** toggle, which defaults to `No`.

If you toggle **Enabled** to `Yes`:

- Set **Validate server certs** to `true` if you want the connection to fail when the TLS server certificate cannot be validated. Defaults to `No`.
- Leave **CA certificate path** blank if you toggled the **Validate server certs** to `true` and are using the local OS-provided trusted CA certificates to validate the server's certificate. Otherwise, specify the server path containing CA certificates, which must be in PEM format.

Logs Settings

Configure where and how AppScope should send its own logs.

Connection type: From the drop-down, choose one of the connection types described above, then enter values in whatever fields the UI exposes.

Log level: Set logging verbosity. When **Send metrics and events over the same connection?** is toggled to `Yes`, **Log level** is locked at `warning`.

File path: Specify a directory in which to store logs. If not specified, defaults to `/tmp/scope.log`.

Buffering: Set this method to `Line` if there's a chance that multiple scoped processes will be writing to the same file. This prevents the log file from getting scrambled, with interleaved lines. However, in single-writer scenarios, selecting `Full` often improves performance.

Metrics Settings

The Schema Reference in the AppScope docs lists and fully describes every metric that AppScope can emit.

Process (`proc.*`) metrics periodically report information about resource usage at a point in time. All other metrics report on actions taken by a scoped process. For examples, see Metrics and Events in AppScope in the AppScope documentation.

Should AppScope emit metrics?: Defaults to `Yes`, which exposes the following options.

Format: AppScope can emit metrics in either of two formats: `NDJSON` (the default), or `StatsD`. If you choose `StatsD`, the UI exposes the following fields:

- **StatsD prefix**: Optionally, enter a prefix for AppScope to add to metrics.
- **StatsD max length**: Optionally, enter a maximum length, in bytes, for StatsD-formatted metrics. AppScope will truncate any metric which exceeds the maximum length. (Defaults to `512`.)

Verbosity: Set a value between `0` and `9`. (Defaults to `4`.) Verbosity controls both the cardinality of metrics that AppScope emits, and which kinds of metrics AppScope aggregates. For details, see Metrics and Events in AppScope in the AppScope documentation. But the general principle is:

- At low verbosity, metrics summarize multiple actions over a reporting period.
- At higher verbosity, metrics provide details about single actions in real time.

Reporting period: Set the metric summary interval in seconds. Defaults to `10`.

Watch list: Select categories of metrics to emit. Defaults to the following categories: `StatsD`, `File System`, `Network`, `HTTP`, `DNS`, `Process`.

Events Settings

The Schema Reference in the AppScope docs lists and fully describes every event that AppScope can emit.

Should AppsScope emit events?: Defaults to `Yes`, which exposes the following controls in this section.

Event-rate limiter (events per second): AppScope discards events generated in excess of the limiter setting, which defaults to `10000`. To disable the limiter, set it to `0`.

Enhance filesystem event data: When toggled to `Yes` (the default), AppScope enhances `fs.open` events by adding the `uid`, `gid`, and `mode` fields.

Watch list: Specify categories of events to emit, and configure their content and structure. Defaults to the following categories: `Console`, `File`, `Filesystem`, `Network`, `DNS`, and `HTTP`. For details, see Categories.

Events Categories

The **Watch list** is a table providing a row for each enabled category of events. The **Type** column contains the category names. The other columns (**Name**, **Field**, and **Value**) contain regular expressions.

At the right edge of the **Name**, **Field**, and **Value** fields is a Copy button, a Flag button to append Regex flags, and an Expand button to open a validation modal.

AppScope emits an event when its values match all the regexes in the corresponding category's row. For **HTTP**, request and response events must also match the headers listed in the **Header** column.

Allow binary is a toggle, available only for **Console** events. When toggled to `Yes` (the default), **Console** events will include binary data along with text data. If you prefer to redact binary data, toggle to `No`.

The **Type** column enumerates any or all of the following event category names. Toggle **Enabled** to `Yes` to include a category.

- **Console**: Includes writes to standard out and error. Use this to monitor console output, especially in containerized environments where logging to files isn't commonly done.
- **File**: Includes writes to files. Use this to monitor log files and/or any other files of interest.
- **Filesystem**: Includes filesystem operations like `open`, `close`, and `delete`.
- **Network**: Includes `open` and `close` events on network connections.
- **DNS**: Includes DNS request and response events.
- **HTTP**: Includes HTTP request and response events.
- **Granular**: Seldom used, and not emitted by default. Can be useful when tracking down a problem. When this type is enabled, AppScope sends more events than it otherwise would, giving you a more granular view of the activity of the scoped process.

Some usage examples:

- For `fs.*` events, AppScope normally emits `fs.open` and `fs.close` events. With **Granular** enabled, every time a read or write occurs, AppScope emits `fs.read` or `fs.write`, respectively.
- For `net.*` events, AppScope normally emits `net.open`, `net.close`, and `net.app`.With **Granular** enabled, every time a read or write is done on the socket, AppScope emits `net.rx` or `net.tx`, respectively.

Payloads

AppScope never sends encrypted payloads to disk, to Cribl Stream, or to Cribl Edge.

Capture all payloads: Enable capturing all payloads. Defaults to `No`. Enable this setting with caution: when scoping I/O-intensive programs, it can produce large amounts of data.

Protocol detection: Click **Add Protocol** to specify a protocol (e.g., StatsD, HTTP, Redis) for AppScope to detect in network payloads, as well as what AppScope should do upon detecting that protocol. AppScope checks the first packet on a channel against the regular expression for each protocol until it finds a match. Then AppScope treats the matching protocol as the detected protocol for its channel, and ignores any others listed. See Sample Protocols for regexes that detect common protocols.

Name: Enter a short, descriptive name to define the protocol-detect event. AppScope inserts this value into the header for payloads sent to Cribl. Since AppScope sends this as JSON, you cannot include double quotes or backslashes.

Regex: Enter a regular expression to detect the desired protocol.

Generate protocol-detect-events: Toggle to `Yes` to generate protocol-detect events. Defaults to `No`.

Capture payload: Toggle to `Yes` to send payloads from matching streams to the configured destinations. Defaults to `No`.

Convert to hex: When toggled to `Yes` (the default), AppScope converts part of the payload to a hex-string before applying the regex. Otherwise, AppScope applies the regex to the binary payload.

Bytes to convert: The number of bytes to convert to hex before applying the regex. Defaults to `256`.

Payload directory: Specify the directory where AppScope should store payload files. Consider using a performant filesystem to reduce I/O performance impacts. Defaults to `tmp`. When **General Settings** > **Send metrics and events over the same connection?** is toggled to `Yes`, AppScope ignores this field and sends payloads to the default destination.

Advanced Settings

Send process-start message?: When toggled to `Yes` (the default), enables the process-start message. This message is the first one AppScope sends upon establishing a connection. It describes AppScope's configuration for that session, as well as the application being scoped.

Command directory: If you want to dynamically change AppScope configuration while scoping a running process, save a `scope.<pid>` file to a **command directory** of your choice, where `<pid>` is the PID of the scoped process, and the file contains lines of the form `<ENV_VAR>=<value>`. Once per reporting period, AppScope looks for a `scope.<pid>` file in the **command directory**, which defaults to `/tmp`. You can configure reporting period in the **Metrics** tab.

Metadata: If you want to add a field to events and metrics that AppScope emits, click **Add Metadata** and define it as a key-value pair. Key/value definitions cannot include environment variables.

Per-process configs: Optionally, define a config that, given a process that fits a particular pattern, will partly or completely override the config you have defined already.

Click **Add config** and configure the **Filters** relevant to the process(es) you want to match:

- **Process name**: Enter a string to match the basename of the scoped process.
- **Process argument**: Enter an expression to match the scoped process's full command line including options and arguments.
- **Hostname**: Enter a string to match the hostname of the machine where the scoped process is running.
- **Username**: Enter a string to match the username for the scoped process's UID.
- **Environment**: To match when an environment variable is set, enter its name alone (i.e., `FOO`). To match when the variable is set with a particular value, enter both (i.e., `FOO=bar`).
- **Ancestor**: Enter a string to match the basename of the scoped process's parent, parent's parent, etc.

Then, in the **Config** column, click the **Edit** button. This takes you through the settings, beginning with **General Settings**, but this time you'll be specifying what in the previously-defined settings to override when a process matches your **Filters**.

Protocol Detection Examples for Payloads

Here are some example definitions for detecting protocols in payloads.

Redis

This regex detects the plain-text Redis serialization protocol (RESP).

Name: `Redis`

Regex: `"^[*]\\d+|^[+]\\w+|^[$]\\d+"`

Mongo

This regex detects the MongoDB wire protocol. Since it's binary, we convert to hex.

Name: `Mongo`

Regex: `^2401000000000000000000000d407`

Convert to hex: `Yes`

Convert length: `14`

HTTP

By default, AppScope uses an internally-defined protocol detector for HTTP similar to this example.

Name: `HTTP`

Regex: `"HTTP\\/1\\.[0-2]|PRI \\* HTTP\\/2\\.0\r\n\r\nSM\r\n\r\n"`

StatsD

By default, AppScope uses an internally-defined protocol detector for StatsD similar to this example.

Name: `STATSD`

Regex: `"^([^:]+):([\\d.]+)\\|(c|g|ms|s|h)"`

TLS

By default, AppScope uses an internally-defined protocol detector for TLS/SSL similar to this example.

Name: `TLS`

Regex: `"^(?:(?:16030[0-3].{4})|(?:8[0-9a-fA-F]{3}01))"`

Convert to hex: `Yes`

Convert length: `5`

16. Reference

16.1. API Docs

To complement our API Reference, below are some examples of using the Cribl Edge API to address common scenarios. If you want to automate an action in Cribl Edge for which you can't find documentation, ask us in Cribl Community's `#docs` channel.

Breaking Change – v.4.0+

In Cribl Edge 4.0 and later, `PATCH` requests to all APIs – except for `samples` and `system` endpoints – require the resource to already exist. (In previous versions, a `PATCH` request would create a resource that didn't exist.) To accommodate this change, make a `POST` request to create an item before any `PATCH` request to update it.

Using the Correct API URL

Every Cribl API URL includes a server URL and an endpoint path, which vary depending on the type of deployment.

The server URL follows these patterns:

- Cribl.Cloud deployment:
  `https://logstream.<organizationID>.cribl.cloud/<endpoint_path>`
- On-prem deployment:
  `https://<cribl-stream-leader>:9000/<endpoint_path>`

The endpoint path always begins with `/api/v1/`, but the remainder varies. For example, the `/system/outputs` endpoint follows these patterns:

- Single-instance deployment: `/api/v1/system/outputs/{id}`
- Distributed deployment: `/api/v1/m/{workerGroup}/system/outputs/{id}`
- For licenses limited to a single Fleet, the `{workerGroup}` value will always be `default`

When composing requests to the Cribl API, use the pattern that matches your deployment type. Adapt examples from this page in the same way.

In Cribl.Cloud and other distributed deployments, you must **Commit** and **Deploy** your changes after following the steps in the examples. With the API, you can automate commit/deploy commands, too.

Obtaining API Bearer Tokens

Other than calls to `/auth/login` and `/health`, all API requests require a Bearer or access token. You obtain the token in different ways, depending on where your Cribl Edge instance is deployed, as outlined in this section. You can make the listed requests at the command line, programmatically, or using the UI.

On-Prem

To obtain a Bearer token, send a local authentication request and payload to the API as shown below. Adapt the example to include your Leader hostname (or IP address), username, and password.

If you're using SSO/OpenID Connect Authentication, you must toggle **Allow local auth** on, because you'll need to be a Local user when you authenticate via the API.

Request:

```
POST https://<cribl-stream-leader>:9000/api/v1/auth/login
```

Payload:

```
{
    "username": "<username>",
    "password": "<password>"
}
```

Example request using `curl`:

```
curl -X POST https://<cribl-stream-leader>:9000/api/v1/auth/login -H "Content-Type
```

You'll get a JSON object as the response, e.g.:

```
{
    "token": "<JWT_Token>",
    "forcePasswordChange": false
}
```

Use this Bearer token in all subsequent requests. Include it in the Authorization header, like this:

```
Bearer <JWT_token>
```

Cribl.Cloud

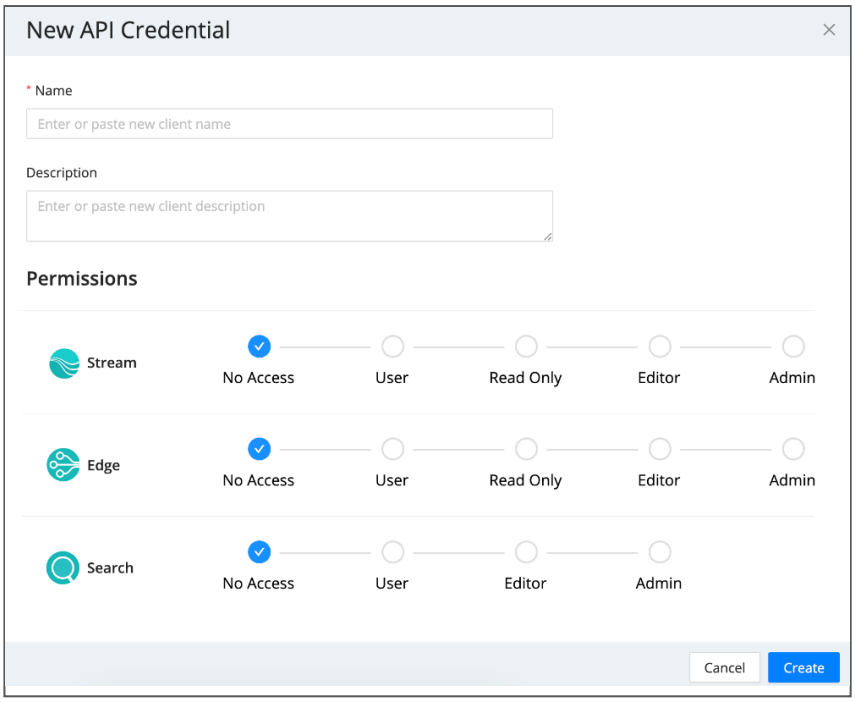
As of Cribl Edge 4.1 (March 2023), Owners and Admins can generate API access tokens directly in their Cribl.Cloud Organization's portal. This is a two-tier process:

- Generate a Credential to expose a Client ID/Client Secret pair.
- Use the Credential's Client ID and Client Secret to obtain an access token, valid for 24 hours, that can be used to run API calls against Cribl APIs.

Creating an API Credential

Owners and Admins can create Credentials. From your Organization's **Options** (•••) menu, first select **Account** > **Organization**. Then:

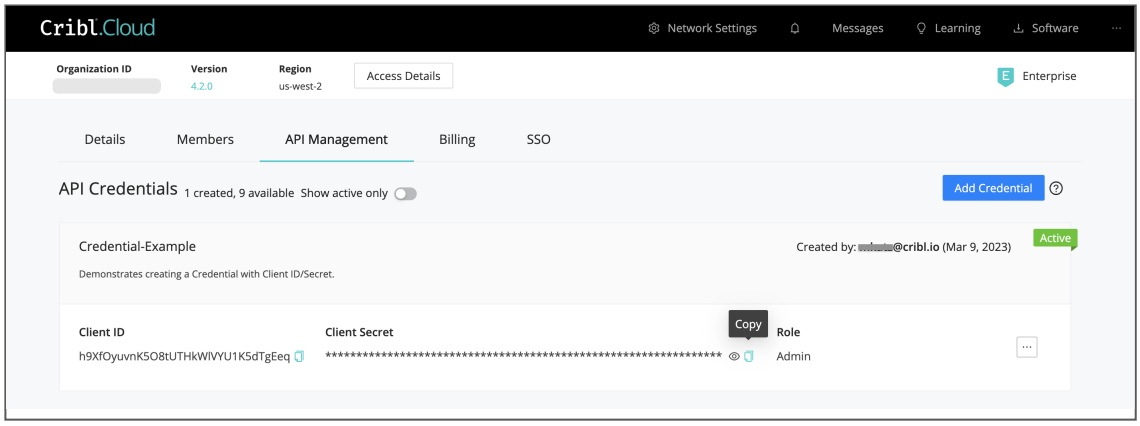
1. Select the **API Management** tab.
2. Click **Add Credential**.
3. Specify a **Name**, an optional **Description** of the Credential's purpose, and the **Permissions** that the Credential's tokens will grant.
   The **Permissions** selector is available with an Enterprise plan. Without an Enterprise plan, all tokens will be granted the Admin Role. See Cribl.Cloud Roles for details on the available permissions.
4. Click **Create** to save the Credential.



Creating a Cribl.Cloud API Credential

Using an API Credential

The new API Credential will appear on the **API Management** page within a few seconds. Each Credential exposes a **Client ID** and **Client Secret**, with Copy buttons. After Credentials are created, Owners and Admins can use them to generate tokens, and then pass these tokens to Cribl APIs.



Getting an API Credential's Client ID and Secret

Obtaining Tokens

Obtain a 24-hour access token using a displayed Credential's Client ID and Client Secret. Example curl request, with placeholders:

```
curl --request POST \
--url "https://login.cribl.cloud/oauth/token" \
--header "content-type: application/json" \
--data '{"grant_type":"client_credentials", "client_id": "<client_id>", "client_se
```

Making API Calls

Using the token returned by the preceding example, you can make requests to Cribl APIs. Example curl request, with placeholder token value:

Single-instance | Distributed

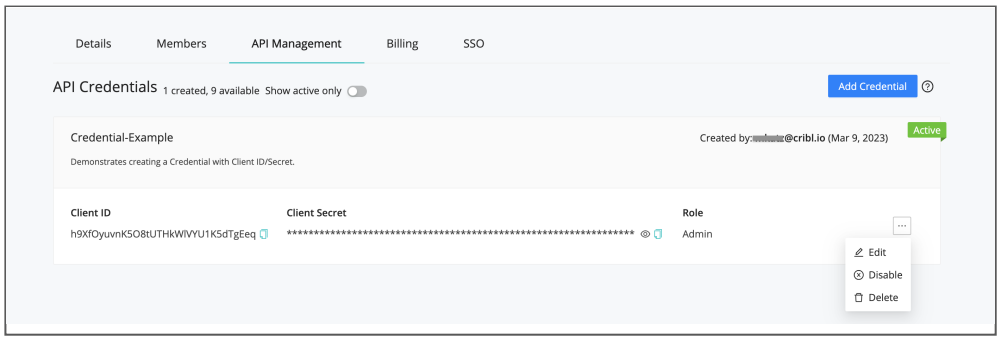
```
curl --location 'https://main-<organizationID>.cribl.cloud/api/v1/m/{workerGrou
--header 'Authorization: Bearer <token-generated-above>'
```

Refreshing Tokens

Admins are responsible for ensuring that their applications obtain a new token within each token's 24-hour expiration window. This can be done at any time before expiration.

Managing API Credentials

A Cribl.Cloud Organization's Owners and Admins can view, edit, and disable existing Credentials. Only Owners can delete Credentials.



Managing Cribl.Cloud access tokens

Update Basic Configurations

You can use the API to programmatically update the configuration of any object type that the API supports (e.g., Sources and Destinations).

Example: Periodically rotate S3 keys on a preconfigured S3/MinIO destination.

1. Send a `GET` request to the `/outputs` endpoint to retrieve the definition for a Destination (in this case, `MinIO`). The response provides the definition in its payload, as a JSON object.
   Example request:

Single-instance | Distributed

```
curl -X GET "<url>/api/v1/m/{workerGroup}/system/inputs/<output id>" -H "accept
```

Example response:

```
{
  "systemFields": [
    "cribl_pipe"
  ],
  "signatureVersion": "v4",
  "objectACL": "private",
  "partitionExpr": "`${host}/${sourcetype}`",
  "format": "json",
  "baseFileName": "CriblOut",
  "compress": "none",
  "maxFileSizeMB": 32,
  "maxFileOpenTimeSec": 300,
  "maxFileIdleTimeSec": 30,
  "maxOpenFiles": 100,
  "onBackpressure": "block",
  "id": "minio",
  "type": "minio",
  "endpoint": "http://minio:9090",
  "bucket": "test",
  "destPath": "keyprefix",
  "stagePath": "tmp",
  "awsApiKey": "minioadmin",
  "awsSecretKey": "minioadmin"
}
```

2. Edit the definition so that when you send it back in a `PATCH` request, it updates the desired Destination's S3 keys.
    - Edit the value of the `id` field to be the ID of the specific Destination whose keys you want to rotate, e.g., `minio_042`.
    - Edit the values of the S3 key fields, e.g., `"awsApiKey": "minioadmin_new_api_key"` and `"awsSecretKey": "minioadmin_new_secret_key"`.
3. Send the edited definition as the payload of a `PATCH` request to the `/outputs/{id}` endpoint. This patches (i.e., updates) the specified MinIO Destination's configuration.

Upload a Lookup File

This section demonstrates how to upload a Lookup file via the API. The following examples assume that we're uploading the file to a Fleet named `default`.

1. Send a `PUT` request to the `/system/lookups` endpoint to upload the file. Example:

Single-instance | Distributed

```
curl -X PUT "<url>/api/v1/m/{workerGroup}/system/lookups?filename=example.csv"
-H "Authorization: Bearer <token>" -H 'Content-Type: text/csv' \
--data-binary '@/path/to/your/example.csv'
```

You will receive a JSON object response similar to the following example:

```
{"filename":"example.csv.random.tmp","rows":100,"size":200}
```

2. Send a `POST` request referencing the Lookup file to the `/system/lookups` endpoint. Replace the filename with the response from the previous `PUT` request. This both creates the Lookup and moves the Lookup file to its final destination. Example:

Single-instance | Distributed

```
curl -X POST "<url>/api/v1/m/{workerGroup}/system/lookups" \
-H "accept: application/json" -H "Authorization: Bearer <token>" \
-H "Content-Type: application/json" \
-d '{"id":"example.csv","fileInfo":{"filename":"example.csv.random.tmp"}}'
```

Example response:

```
{
  "items": [
    {
      "id": "example.csv",
      "size": 200
    }
  ],
  "count": 1
}
```

3. If replacing an existing lookup, send a `PATCH` request referencing the existing filename in the URL and the body. Example:

Single-instance | Distributed

```
curl -X PATCH "<url>/api/v1/m/{workerGroup}/system/lookups/example.csv" \
-H "accept: application/json" -H "Authorization: Bearer <token>" \
-H "Content-Type: application/json" \
-d '{"id":"example.csv","fileInfo":{"filename":"example.csv.random.tmp"}}'
```

Example response:

```
{
  "items": [
    {
      "id": "example.csv",
      "size": 200
    }
  ],
  "count": 1
}
```

Distributed Upload

In a distributed environment, for lookup files of manageable size: Cribl recommends uploading the Lookup file only to the Leader Node, and then making a selective commit and deploy with only those Lookup file changes. The Leader then notifies Edge Nodes that a new configuration is available, and Edge Nodes will pull the new configuration from the Leader Node.

Creating HEC Tokens with Python

Cribl SE Jon Rust has written a Python script which demonstrates how to authenticate to the Cribl API, make a simple POST request, and add a new HEC token.

To use the script, you'll need:

- Python 3.
- The Python 3 Requests module (use brew or pip3 to install).
- A working, distributed Cribl Edge installation, with a configured Splunk HEC Source.
- An admin username and password.

The script and instructions for usage can be found in Jon Rust's GitHub repo.

Managing Packs via APIs

You can perform Pack operations by running Cribl API calls on the command line. This is required if you plan to automate Pack operations – e.g., in a CI/CD pipeline. For more details, see Managing Packs via API .

16.2. CLI Reference

Command line interface basics

In addition to starting and stopping the Cribl Edge server, Cribl Edge's command line interface enables you to initiate many configuration and administrative tasks directly from your terminal.

Command Syntax

To execute CLI commands, the basic syntax is:

```
cd $CRIBL_HOME/bin
./cribl <command> <sub-command> <options> <arguments>
```

Not all commands have sub-commands.

To see help for any command, append the `--help` option, for example:

```
./cribl vars --help

./cribl vars get --help

./cribl vars get -i myArray --help
```

The `scope` command is an exception: it has no `--help` option, but it has its own CLI Reference in the AppScope documentation.

Avoiding Surprises

Immediate Execution

As indicated in the sample output below, some commands take effect immediately.

Commands that require further input will echo the sub-commands, options, and arguments they expect.

Persistent Volumes

If you start Cribl Edge with the `CRIBL_VOLUME_DIR` variable, all subsequent CLI commands should have this variable defined. Otherwise, those commands will apply Cribl Edge's default directories, yielding misleading results.

You can set `CRIBL_VOLUME_DIR` as an environment variable, or you can explicitly include it in each command, as in this example:

```
CRIBL_VOLUME_DIR=<writable-path-name> /opt/cribl/bin/cribl status
```

When set, `$CRIBL_VOLUME_DIR` overrides `$CRIBL_HOME`.

Avoid setting `$CRIBL_VOLUME_DIR` to an existing folder because it creates predefined folders in the specified directory. If that directory already contains folders with those names, they will be overwritten.

Commands Available

To see a list of available commands, enter `./cribl` alone (or the equivalent `./cribl help`). To execute a command, or to see its required parameters, enter `./cribl <command>`.

`help`

Displays a list of commands with a description (help) for each. Defaults to a selection of generally useful commands.

Usage

```
./cribl help [-a]
```

Options

```
  -a              - Display the list of all commands, except for `scope`.
```

Sample Response

```
Software version: 4.4.2
Usage: [sub-command] [options] [args]

Commands:
help               - Display help
mode-edge          - Configure instance in Edge mode
mode-managed-edge - Configure instance in Managed-Edge mode
mode-master        - Configure instance in Leader mode
mode-single        - Configure instance in Single-Instance mode
mode-worker        - Configure instance in Worker mode
reload             - Reload instance
restart            - Restart instance
start              - Start instance
status             - Display status
stop               - Stop instance
version            - Print version

auth               - Authentication
boot-start         - Enable/Disable boot-start
diag               - Manage diagnostics bundles
git                - Manage Worker Groups config
keys               - Manage encryption keys
nc                 - Listen on a port for traffic and output stats and data
node               - Execute a JavaScript file
pack               - Manage Cribl Packs
parquet            - Manage Parquet files and schemas
pipe               - Feed stdin to a pipeline
vars               - Manage global variables
```

Starting in version 3.0, Cribl Edge's former "master" application components have been renamed "Leader." As long as some legacy terminology remains within CLI commands/options, configuration keys/values, and environment variables, this document will reflect the options available.

mode-master

Configures Cribl Edge as a Leader instance.

Usage

```
./cribl mode-master <options> <args>
```

Options

```
  [-H <host>]             - Host (defaults to 0.0.0.0).
  [-p <port>]             - Port (defaults to 4200).
  [-n <certName>]         – Name of saved certificate. Mutually exclusive with -k or
  [-k <privKeyPath>]      – Path on server to the private key to use. PEM format. Ca
  [-c <certPath>]         – Path on server to the certificate to use. PEM format. Ca
  [-u <authToken>]        - Optional authentication token to include as part of the
  [-i <ipWhitelistRegex>] – Regex matching IP addresses that are allowed to establis
  [-r <resiliency>]       – Resiliency mode for the Leader. Accepts agruments: none,
  [-v <failoverVolume>]   – If the -r option is set to failover, this defines the vo
```

Sample Response

```
  Settings updated.
  You will need to restart Cribl Stream before your changes take full effect.
```

mode-single

Configures Cribl Stream as a single-instance deployment.

Usage

```
./cribl mode-single [--help]
```

Sample Response

```
  Settings updated.
  You will need to restart Cribl Stream before your changes take full effect.
```

mode-edge

Configures Cribl Edge as a single-instance deployment.

Usage

```
./cribl mode-edge [--help]
```

Options

```
  [-H <host>]   - Hostname/IP (defaults to 127.0.0.1).
  [-p <port>]   - Port (defaults to 9420).
  [-s <socket>] - Location of the AppScope socket.
```

Sample Response

```
  Settings updated.
```

`mode-worker`

Configures Cribl Stream as a Worker instance.

Usage

```
./cribl mode-worker -H <host> -p <port> <options> <args>
```

The `-H <host> -p <port>` parameters are required.

Options

```
-H <host>           - Leader Node's Hostname or IP address.
-p <port>           - Leader Node's cluster communications port (defaults to 4200).
[-S <true|false>]   - Sets, or disables, secure communication between Leader and Wo
[-n <certName>]     - Name of saved certificate. Mutually exclusive with -k or -c.
[-k <privKeyPath>] - Path on server to the private key to use. PEM format. Can refe
[-c <certPath>]     - Path on server to the certificate to use. PEM format. Can refe
[-u <authToken>]    - Authentication token to include as part of the connection hea
[-e <envRegex>]     - Regex that selects environment variables to report to Leader.
[-t <tags>]         - Tag values to report to Leader.
[-g <group>]        - Worker Group/Fleet to report to Leader.
```

Sample Response

```
Settings updated.
You will need to restart Cribl Stream before your changes take full effect.
```

When you generate bootstrap scripts via the UI to add or update Workers, these scripts automatically set the `-S` option according to the Leader's TLS configuration.

mode-managed-edge

Configures Cribl Edge as an Edge Node.

Usage

```
./cribl mode-managed-edge -H <host> -p <port> <options> <args>
```

The `-H <host> -p <port>` parameters are required.

Options

```
-H <host>           - Leader Node's Hostname or IP address.
-p <port>           - Leader Node's cluster communications port (defaults to 4200).
[-S <true|false>]   - Sets, or disables, secure communication between Leader and Wo
[-n <certName>]     - Name of saved certificate. Mutually exclusive with -k or -c.
[-k <privKeyPath>] - Path on server to the private key to use. PEM format. Can refe
[-c <certPath>]     - Path on server to the certificate to use. PEM format. Can refe
[-u <authToken>]    - Authentication token to include as part of the connection hea
[-e <envRegex>]     - Regex that selects environment variables to report to Leader.
[-t <tags>]         - Tag values to report to Leader.
[-g <group>]        - Worker Group/Fleet to report to Leader.
```

Sample Response

```
Settings updated.
```

When you generate bootstrap scripts via the UI to add or update Workers, these scripts automatically set the `-S` option according to the Leader's TLS configuration.

pack

Manages Cribl Packs.

Usage

```
./cribl pack <sub-command> <options> <args>
```

Sub-commands and Options

```
export            - Export Cribl Packs, args:
   -m <mode>      - Mode to export. Accepts: merge_safe, merge, default_only.
  [-o <filename>] - Where to export the pack on disk.
  [-n <name>]     - Name to override the installed pack's name on export.
  [-g <group>]    - The Worker Group/Fleet to execute within
install           - Install a Cribl Pack, args:
  [-d]            - Run install in debug.
  [-f]            - Force install.
  [-c]            - Disallow installation of Packs with custom functions.
  [-n <name>]     - Name of the pack to install; defaults to source.
  [-g <group>]    - The Worker Group to execute within
list              - List Cribl Packs, args:
  [-v]            - Display all pack info
  [-g <group>]    - The Worker Group to execute within
uninstall         - Uninstall a Cribl Pack, args:
  [-d]            - Run uninstall in debug
  [-g <group>]    - The Worker Group to execute within
upgrade           - Upgrade a Cribl Pack, args:
  [-d]            - Run upgrade in debug
  [-s <source>]   - Provide the pack source
  [-m <minor>]    - Only upgrade to minor version
  [-g <group>]    - The Worker Group to execute within
```

Sample Response

```
id          version  spec  displayName    author       description
-------------------------------------------------------------------------------
HelloPacks  1.0.0    ----  Hello, Packs!  Cribl, Inc.  A sample pack with a simple
```

`parquet`

For any Parquet file: view (`cat`) the file, its metadata (`details`), or its `schema`. This command is available only on Linux.

Usage

```
./cribl parquet <sub-command> -f <file>
```

Sub-commands and Options

```
cat             - View the file, args:
  -f <file>     - Path to file
  [-m <max>]    - Maximum number of events to print
details         - View the file's metadata, args:
  -f <file>     - Path to file
schema          - View the file's Parquet schema, args:
  -f <file>     - Path to file
```

Examples

View a Parquet file:

```
./cribl parquet cat -f /some_parquet_file
```

```
{"__bytes":274,"Unnamed: 0":0,"name":"apples","quantity":10,"price":2.6,"date":"20
...
```

View a Parquet file's metadata, including encoding and compression:

```
./cribl parquet details -f /some_parquet_file
```

```
{
  "schema": {
    "Unnamed: 0": {
      "optional": true,
      "type": "INT64"
    },
    ...
  },
  "path": "../../../src/sluice/js/util/__tests__/data/parquet/single/small_fruits.p
  "num_rows": 40000,
  "num_cols": 11,
  "num_real_cols": 11,
  "num_row_groups": 1,
  "created_by": "parquet-cpp-arrow version 11.0.0",
  "schema_root_name": "schema",
  "metadata": {
    "pandas": "{\"index_columns\": [{\"kind\": \"range\", \"name\": null, \"start\"
    ...
  },
  "version": "PARQUET_2_6",
  "metadataSize": 6072,
  ...
    }
  ]
}
```


View a Parquet file's schema (not including encoding or compression), expressed as JSON like in the Parquet Schema editor:

```
./cribl parquet schema -f /some_parquet_file
```

```
{
  "Unnamed: 0": {
    "optional": true,
    "type": "INT64"
  },
  "name": {
    "optional": true,
    "type": "STRING"
  },
  "quantity": {
    "optional": true,
    "type": "DOUBLE"
  },
  "price": {
    "optional": true,
    "type": "DOUBLE"
...
  }
}
```


reload

Reloads Cribl Edge. Executes immediately.

Usage

```
./cribl reload [--help]
```

Sample Response

```
Reload request submitted to Cribl
```

restart

Restarts Cribl Edge. Executes immediately.

Executing this command cancels any running collection jobs.

Usage

```
./cribl restart [--help]
```

Sample Response

```
Stopping Cribl, process 18
............
Cribl Stream is not running
Starting Cribl Stream...
...
Cribl Stream started
```

start

Starts Cribl Edge. Executes immediately. Upon first run, echoes Cribl Edge's default login credentials.

Usage

```
./cribl start <options> <args>
```

Options

```
[-d <dir>]  - Configuration directory
[-r <role>] - Process role
```

Sample Response

```
Starting Cribl Stream...
...
Cribl Stream started
```

status

Displays status of Cribl Edge, including the API Server address, instance's mode (Leader or Worker), process ID, and GUID (fictitious example below). Executes immediately.

Usage

```
./cribl status [--help]
```

Sample Response

```
Cribl Stream Status

Address: http://172.17.0.3:9000
Mode: master
Status: Up
Software Version: 3.1.0-f765e418
Config Version: 347079c
Master: 0.0.0.0:4200
PID: 4100
GUID: e706052a-ace9-4511-a7c7-b58a414a07d3
```

stop

Stops Cribl Edge. Executes immediately.

Executing this command cancels any running collection jobs.

Usage

```
./cribl stop [--help]
```

Sample Response

```
Stopping Cribl Stream, process 3951
.............
Cribl Stream is not running
```

version

Displays Cribl Edge version. Executes immediately.

Usage

```
./cribl version [--help]
```

Sample Response

```
Software Version: 3.1.0-f765e418
```

auth

Log into or out of Cribl Edge.

Usage

```
./cribl auth <sub-command> <options> <args>
```

Sub-commands and Options

```
login             - Log in to Cribl Stream/Edge, args:
  [-H <host>]     - Host URL (e.g. http://localhost:9000)
  [-u <username>] - Username
  [-p <password>] - Password
  [-f <file>]     - File with credentials
logout            - Log out from Cribl Stream/Edge
```

Login Examples

Launch interactive login:

```
$CRIBL_HOME/bin/cribl auth login
```

Append credentials as command arguments:

```
$CRIBL_HOME/bin/cribl auth login -h <url> -u <username> -p <password>
```

All `-h` and `host` arguments are optional, provided that the API host and port are listed in the `cribl.yml` file's `api:` section.

Provide credentials in environment variables:

```
CRIBL_HOST=<url> CRIBL_USERNAME=<username> CRIBL_PASSWORD=<password>
$CRIBL_HOME/bin/cribl auth login
```

Provide credentials in a file:

```
$CRIBL_HOME/bin/cribl auth login -f <path/to/file>
```

Corresponding file contents:

```
host=<url>
username=<username>
password=<password>
```

boot-start

Enables or disables Cribl Edge boot-start.

Usage

```
./cribl boot-start <sub-command> <options> <args>
```

Sub-commands and Options

```
disable             - Disable Cribl Stream/Edge boot-start, args:
  [-m <manager>]    - Init manager (systemd|initd)
  [-c <configDir>]  - Config directory for the init manager
enable              - Enable Cribl Stream/Edge boot-start, args:
  [-m <manager>]    - Init manager (systemd|initd)
  [-u <user>]       - User to run Cribl Stream/Edge as
  [-c <configDir>]  - Config directory for the init manager
```

Sample Response

```
Enabling Cribl Stream/Edge to be managed by initd...
boot-start enable command needs root privileges...
Enabled Cribl Stream/Edge to be managed by initd as user=root.
```

diag

Manages diagnostic bundles.

Usage

```
./cribl diag <sub-command> <options> <args>
```

Sub-commands and Options

```
  create                    - Creates diagnostic bundle for Cribl Stream/Edge, args:
    [-d]                    - Run create in debug mode
    [-j]                    - Do not append '.txt' to js files
    [-t <maxIncludeJobs>] - Latest number of jobs to include in bundle
    [-M]                    - Exclude metrics from bundle
    [-g]                    - Exclude git log from bundle
  list                      - List existing Cribl Stream/Edge diagnostic bundles
  send                      - Send Cribl Stream/Edge diagnostics bundle to Cribl Suppo
     -c <caseNumber>        - Cribl Support Case Number
    [-p <path>]             - Diagnostic bundle path (if empty then new bundle will be
  heapsnapshot              - Generate heap snapshot of a Cribl Stream/Edge process, a
    [-p <pid>]              - The pid of the process to dump the heap snapshot
```

Sample Responses

`diag create`

```
Created a Cribl diagnostic bundle at /opt/cribl/diag/<product>-zedborcdb72f-202108
```

`diag heapsnapshot -p 12345`

The response format is `Heap-<epoch-timestamp>-<pid>.heapsnapshot`:

```
Heap-1672574400000-12345.heapsnapshot
```

`git`

Manages Worker Group/Fleets'/Fleets' configuration.

Usage

```
./cribl git <sub-command> <options> <args>
```

Sub-commands and Options

```
commit           - Commit, args:
  [-g <group>]   - Group ID.
  [-m <message>] - Commit message.
commit-deploy    - Commit & Deploy, args:
   -g <group>    - Group ID.
  [-m <message>] - Commit message.
deploy           - Deploy, args:
   -g <group>    - Group ID.
  [-v <version>] - Deploy version.
list-groups      - List Worker Groups/Fleets.
```

Sample Response

```
Successfully committed version 7c04de1
```

`groups`

Deprecated. See `git`.

`keys`

Manages encryption keys. You must append the `-g <group>` argument to specify a Worker Group/Fleet. As a fallback, append the argument `-g default`, e.g.: `./cribl keys list -g default`

Usage

```
./cribl keys <sub-command> <options> <args> -g <group>
```

Sub-commands and Options

```
add                - Add encryption keys, args:
  [-a <algorithm>] - Encryption algorithm. Supported values: aes-256-cbc (default),
  [-c <keyclass>]  - Key class to set for the key.
  [-k <kms>]       - KMS to use. Must be configured; see cribl.yml.
  [-e <expires>]   - Expiration time, in epoch time.
  [-i]             - Use an initialization vector. (IV size configurable for algor
  [-s <ivSize>]    - (For algo 'aes-256-gcm' only) Initialization vector size (byt
  [-g <group>]     - Worker Group's ID.
list               - List encryption keys, args:
  [-g <group>]     - Worker Group's ID.
```

Sample Response

```
Adding key succeeded. Key count=1
```

nc

Listens on a port for traffic, and outputs stats and data. (Netcat-like utility.)

Usage

```
./cribl nc -p <port> <options> <args>
```

Options

```
 -p <port>           - Port to listen on.
[-f <family>]        - If this is "6" then nc will listen on ::1, otherwise it list
[-s <statsInterval>] - Stats output interval (ms), use 0 to disable.
[-u]                 - Listen on UDP port instead.
[-o]                 - Output received data to stdout.
[-t <throttle>]      - throttle rate in (unit)/sec, where units can be KB,MB,GB, ar
```

Sample Response

```
2021-08-20T22:44:30.457Z - starting server on 0.0.0.0:9999
2021-08-20T22:44:30.462Z - server listening 0.0.0.0:9999
2021-08-20T22:44:31.461Z - messages: 0, socks: 0, thruput: 0MBps
2021-08-20T22:44:32.466Z - messages: 0, socks: 0, thruput: 0MBps
...
2021-08-20T22:44:39.212Z - got connection: 127.0.0.1:37190
2021-08-20T22:44:39.213Z - got connection: 127.0.0.1:37192
```

node

Run with no options, displays a command prompt, as shown here:

```
>
```

To execute a JavaScript file, you can enter path/filename at the prompt.

With the `-v` option, prints the version of NodeJS that is running.

With `-e`, evaluates a string. Write to console to see the output, for example:

```
./cribl node -e 'console.log(Date.now())'
1629740667695
```

Usage

```
./cribl node <options> <args>
```

Options

```
[-e <eval>] - String to eval
[-v]        - Prints NodeJS version
```

Sample Response

```
v14.15.1
```

pipe

Feeds stdin to a pipeline.

Usage

```
./cribl pipe -p <pipelineName> <options> <args>
```

Examples:

```
cat sample.log |  ./cribl pipe -p <pipelineName>
cat sample.log |  ./cribl pipe -p <pipelineName> 2>/dev/null
```

Options

```
 -p <pipeline>    - Pipeline to feed data thru
[-d]              - Include dropped events
[-c <cpuProfile>] - Perform CPU profiling
[-t]              - Perform pipeline tracing
[-a <pack>]       - Optional Cribl Pack context. Mutually exclusive with -b.
[-b <project>]    - Optional Cribl Project context
[-i]              - Runs the pipe command without sandboxing javascript expressions
```

Sample Response

```
...
{"time":"2021-08-20T20:37:00.017Z","cid":"api","channel":"commands","level":"info",
{"time":"2021-08-20T20:37:00.019Z","cid":"api","channel":"pipe:main","level":"info"
{"time":"2021-08-20T20:37:00.021Z","cid":"api","channel":"pipe:main","level":"info"
{"time":"2021-08-20T20:37:00.022Z","cid":"api","channel":"commands","level":"info",
{"time":"2021-08-20T20:37:00.028Z","cid":"api","channel":"GrokMgr","level":"info","
...
```

scope

Greps your apps by the syscalls. Executes immediately.

See the AppScope CLI Reference for usage and examples.

vars

Manages Cribl Edge Global Variables.

Usage

```
./cribl vars <sub-command> <options> <args>
```

Sub-commands and Options

```
Sub-commands:
add                   - Add global variable, args:
   -i <id>            - Global variable ID
   -t <type>          - Type
   -v <value>         - Value
  [-a <args>]         - Arguments
  [-d <description>] - Description
  [-c <tags>]         - Custom Tags (comma separated list)
  [-g <group>]        - Group ID
get                   - List global variables, args:
  [-i <id>]           - Global variable ID
  [-g <group>]        - Group ID
remove                - Remove global variable, args:
   -i <id>            - Global variable ID
  [-g <group>]        - Group ID
update                - Update global variable, args:
   -i <id>            - Global variable ID
  [-t <type>]         - Type
  [-v <value>]        - Value
  [-a <args>]         - Arguments
  [-d <description>] - Description
  [-c <tags>]         - Custom Tags (comma separated list)
  [-g <group>]        - Group ID
```

Sample Response


```
[
  {
    "type": "number",
    "lib": "cribl",
    "description": "Sample number variable ",
    "value": "42",
    "tags": "cribl,sample",
    "id": "theAnswer"
  }
]
```

16.3. Config Files

Understanding Configuration Paths and Files

Even though all Cribl Edge Routes, Pipelines, and Functions can be managed from the UI, it's important to understand how the configuration works under the hood. Here is how configuration paths and files are laid out on the filesystem.

Path Placeholder	Expanded Path
`$CRIBL_HOME`	Standalone Install: `/path/to/install/cribl/` – referred to below as `$CRIBL_HOME` Cribl App for Splunk Install: `$SPLUNK_HOME/etc/apps/cribl/`

All paths below are relative to `$CRIBL_HOME` in a single-instance deployment, or to `$CRIBL_HOME/groups/<group-name>/` in a distributed deployment.

Category	Relative Path
Default Configurations Out-of-the-box defaults (rewritable) and libraries (expandable)	`default/cribl`
Local Configurations User-created integrations and resources	`local/cribl`
System Configuration	`(default|local)/cribl/cribl.yml` See cribl.yml
API Configuration	`(default|local)/cribl/cribl.yml` > `[api]` section See cribl.yml
Source Configuration	`(default|local)/cribl/inputs.yml` See inputs.yml
Destination Configuration	`(default|local)/cribl/outputs.yml` See outputs.yml
License Configuration	`(default|local)/cribl/licenses.yml`

Category	Relative Path
Regexes Configuration	`(default|local)/cribl/regexes.yml`
Breakers Configuration	`(default|local)/cribl/breakers.yml`
Limits Configuration	`(default|local)/cribl/limits.yml`
Service Processes Configuration	`(default|local)/cribl/service.yml` See service.yml
Pipelines Configuration	`(default|local)/cribl/pipelines/<pipeline_name>` Each Pipeline's config resides within its subdirectory.
Packs Configuration	`default/<pack_name>` Each Pack's code and config reside within its subdirectory.
Routes Configuration	`(default|local)/cribl/pipelines/routes.yml`
Functions	`(default|local)/cribl/functions/<function_name>` Each Function's code resides within its subdirectory.
Functions Configuration	`(default|local)/cribl/functions/<function_name>/...` Each Function's config resides within its subdirectory.
Roles Configuration	`(default|local)/cribl/roles.yml` RBAC Role definitions. See roles.yml.
Policies Configuration	`(default|local)/cribl/policies.yml` RBAC Policy definitions. See policies.yml.
Permissions Configuration	`(default|local)/cribl/perms.yml` User permissions. See perms.yml.
Secrets Configuration	`(default|local)/cribl/secrets.yml` Cribl Edge secrets. See secrets.yml.

Configurations and Restart

You can **Restart** and **Reload** via the UI. On the top nav, go to **Settings** > **Global Settings** > **System** > **Controls** then click on the **Reload** button or the **Restart** button.

In a distributed environment, Edge Nodes poll the Leader for configuration changes. Many of these changes require a quick reload to read the new configuration, while others require a restart of the Cribl processes on the Edge Node.

Upon restarts, be aware of the following:

- Syslog data still being received over UDP might be dropped.
- Edge Nodes will temporarily disappear from the Leader's **Manage Workers** or **Manage Edge Nodes** page.
- Aggregation and suppression operations will start over.
- Edge Nodes' local copies of Monitoring metrics will be dropped.
- Cribl Edge will drop any events still in RAM that were bound for persistent queues. (However, PQ data already written to disk will persist through the restart.)

Changes that require reloads include configuration changes to:

- Functions
- Pipelines
- Packs
- Routes
- Lookups
- Parquet schemas
- Global variables
- Fleet Settings > Limits
- Fleet Settings > Logging > Levels

Changes that require restarts include configuration changes to:

- Distributed mode (Leader versus Managed Edge Node or Single instance)
- Fleet assignment
- Event Breakers
- QuickConnect configs
- Sources
- Destinations
- Fleet Settings > General Settings > TLS
- Fleet Settings > General Settings > Advanced
- Fleet Settings > Worker Processes > Process count and Memory

Some general guidelines to keep in mind:

- Configuration changes generated by most UI interactions – for instance, changing the order of Functions in a Pipeline, or changing the order of Routes – **do not require restarts**.

- Some configuration changes in the **Settings** UI **do require restarts**. These will prompt you for confirmation before restarting.
- All direct edits of configuration files in `(bin|local|default)/cribl/...` **will require restarts**.
- Edge Nodes might temporarily disappear from the Leader's **Workers** or **Edge Nodes** tab while restarting.
- A `git commit` command on the Leader Node's host (using a freestanding `git` client not embedded in Cribl's CLI or UI) **will require either a reload or restart**.
- When using the Cribl App for Splunk, changes to Splunk configuration files might or might not require restarts. Please check current Splunk docs.

Configuration Layering and Precedence

As on most *nix systems, Cribl configurations in `local` take precedence over those in `default`. There is no layering of configuration files.

Editing Configuration Files Manually

When config files **must** be edited manually, save all changes in `local`.

16.3.1. Cribl.Yml

`cribl.yml` contains settings for configuring API and other system properties.

$CRIBL_HOME/default/cribl/cribl.yml

```
api: # [object]
  host: # [string] Address to bind to. Defaults to 0.0.0.0.
  port: # [number] Port to listen to. Defaults to 9000.
  retryCount: # [number] Number of times to retry binding to API port. Defaults to
  retrySleepSecs: # [number] Period, in seconds, between consecutive port binding r
  baseUrl: # [string] URL base path from which to serve all assets (useful when beh
  disabled: # [boolean] Flag to enable/disable local UI access. Defaults to false.
  workerRemoteAccess: # [boolean] Enable teleporting to Edge Nodes'. Defaults to fa
  revokeOnRoleChange: # [boolean] Log users out when their roles change. Defaults t
  authTokenTTL: # [number] How long (in seconds) authentication tokens remain valid
  idleSessionTTL: # [number] How long (in seconds) Cribl Edge will observe no user
  headers: # [object] Custom HTTP headers to be sent with every response.
  loginRateLimit: # [string] Rate limit, expressed as maximum number of requests pe
  ssoRateLimit: # [string] Rate limit for SSO and SLO callback endpoints. Expressed
  apiCache: # [object] Toggle to 'No' to disable caching of browser's frequent API
  ssl: # [object] SSL Settings
    disabled: # [boolean] SSL is enabled by default.
    privKeyPath: # [string] Path to private key.
    certPath: # [string] Path to certificate.

auth: # [object] Authentication Settings
  type: # [string] Type - Select one of the supported authentication providers.

  # -------------- if type is ldap ---------------

  secure: # [boolean] Secure - Enable to use a secure ldap connection (ldaps://), d
  ldapServers: # [array of strings] LDAP servers - List of LDAP servers, each entry
  bindDN: # [string] Bind DN - Distinguished name of entity to authenticate with LD
  bindCredentials: # [string] Password - Distinguished Name password used to authen
  searchBase: # [string] User search base - Starting point to search LDAP for users
  usernameField: # [string] Username field - LDAP user search field, e.g. cn or uid
  searchFilter: # [string] User search filter - LDAP search filter to apply when fi
  groupSearchBase: # [string] Group search base - Starting point to search LDAP for
  groupMemberField: # [string] Group member field - LDAP group search field, e.g. m
  groupSearchFilter: # [string] Group search filter - LDAP search filter to apply w
  groupField: # [string] Group name field - LDAP group field, e.g. cn
  connectTimeout: # [number] Connection timeout (ms)
  rejectUnauthorized: # [boolean] Reject unauthorized - Valid for secure LDAP conne
  fallback: # [boolean] Fallback on fatal error - Attempt local authentication if L
  groups: # [object]
    default: # [string] Default role - Default role assigned to groups not explicit
    mapping: # [object] Mapping - Group(s)-to-role(s) mappings
```

```
# ------------------------------------------------------


# -------------- if type is saas ---------------

issuer: # [string] Issuer - Issuer from which to accept and validate JWT tokens.
tenantId: # [string] Organization ID - The organization ID within which this inst
loginUrl: # [string] Login URL - The URL to redirect unauthenticated users to.

# ------------------------------------------------------


# -------------- if type is splunk ---------------

host: # [string] Host - Hostname or address of Splunk instance that provides auth
port: # [number] Port - Port of Splunk instance that provides authentication. Def
ssl: # [boolean] SSL - Whether SSL is enabled on Splunk instance that provides au
fallback: # [boolean] Fallback on fatal error - Attempt local authentication if S
groups: # [object]
  default: # [string] Default role - Default role assigned to groups not explicit
  mapping: # [object] Mapping - Group(s)-to-role(s) mappings

# ------------------------------------------------------


# -------------- if type is openid ---------------

name: # [string] Provider name - The name of the identity provider service. Manua
audience: # [string] Audience - The Audience from provider configuration. This wi
client_id: # [string] Client ID - The client_id from provider configuration.
client_secret: # [string] Client secret - The client_secret provider configuratio
scope: # [string] Scope - Space-separated list of authentication scopes. Default:
auth_url: # [string] Authentication URL - The full path to the provider's authent
token_url: # [string] Token URL - The full path to the provider's access token UR
userinfo_url: # [string] User Info URL - The full path to the provider's user int
logout_url: # [string] Logout URL - The full path to the provider's logout URL. I
userIdExpr: # [string] User identifier - Expression used to derive userId from th
rejectUnauthorized: # [boolean] Validate certs - Validate certificates; set to fa
filter_type: # [string] Filter type - Optional method for limiting access per use
groupField: # [string] Group name field - Field on the id_token that contains the
fallback: # [boolean] Allow local auth - Allows locally configured users to log i
groups: # [object]
```

```
    default: # [string] Default role - Default role assigned to groups not explici
    mapping: # [object] Mapping - Group(s)-to-role(s) mappings

  # ----------------------------------------------------------

system: # [object]
  upgrade: # [string]
  restart: # [string]
  installType: # [string]
workers: # [object]
  count: # [number]
  memory: # [number]
tls: # [object] Default TLS Settings
  minVersion: # [string] Minimum TLS version - Minimum TLS version. Defaults to TLS
  maxVersion: # [string] Maximum TLS version - Maximum TLS version. Defaults to TLS
  defaultCipherList: # [string] Default cipher list - Default suite of enabled and
      ECDHE-RSA-AES128-GCM-SHA256:
      ECDHE-ECDSA-AES128-GCM-SHA256:
      ECDHE-RSA-AES256-GCM-SHA384:
      ECDHE-ECDSA-AES256-GCM-SHA384:
      DHE-RSA-AES128-GCM-SHA256:
      ECDHE-RSA-AES128-SHA256:
      DHE-RSA-AES128-SHA256:
      ECDHE-RSA-AES256-SHA384:
      DHE-RSA-AES256-SHA384:
      ECDHE-RSA-AES256-SHA256:
      DHE-RSA-AES256-SHA256:
      HIGH:
      !aNULL:
      !eNULL:
      !EXPORT:
      !DES:
      !RC4:
      !MD5:
      !PSK:
      !SRP:
      !CAMELLIA
  defaultEcdhCurve: # [string] ECDH curve - The curve name, or a colon-separated li
  rejectUnauthorized: # [boolean] Validate server certs - Whether to validate serve
proxy: # [object]
  useEnvVars: # [boolean]
git: # [object]
  branch: # [string] Branch - The branch to track in your Stream deployment's git r
```

```
    gitOps: # [string] GitOps workflow - The GitOps workflow for managing Stream's co
    commitDeploySingleAction: # [boolean] Collapse actions - When enabled, Commit & D
    defaultCommitMessage: # [string] Default commit message - Enter a default message
    remote: # [string] Remote URL - Enter remote git repo's URL.
    authType: # [string] Authentication type - Git authentication type.

    # -------------- if authType is ssh ---------------

    sshKey: # [string] SSH private key - Enter SSH private key (without passphrase) t
    strictHostKeyChecking: # [boolean] SSH strict host key checking - Validate key ag

    # --------------------------------------------------------


    # -------------- if authType is basic ---------------

    user: # [string] User - Username for authentication.
    password: # [string] Password - Password for authentication. (With GitHub, use a

    # --------------------------------------------------------

    autoAction: # [string] Scheduled global actions - Global git actions to run autor

    # -------------- if autoAction is commit ---------------

    autoActionSchedule: # [string] Schedule - Cron schedule to run selected git actio
    autoActionMessage: # [string] Commit message - Default scheduled commit message.

    # --------------------------------------------------------


    # -------------- if autoAction is push ---------------

    autoActionSchedule: # [string] Schedule - Cron schedule to run selected git actio

    # --------------------------------------------------------

    timeout: # [number] Git timeout - Max time (milliseconds) to wait for git process
```

Example `cribl.yml`:

$CRIBL_HOME/default/cribl/cribl.yml

```
api:
  host: 0.0.0.0
  port: 9000
  retryCount: 120
  retrySleepSecs: 5
  baseUrl: ""
  # Flag to enable/disable UI. Default: false
  disabled : false
  loginRateLimit: 2/second
  ssl:
    disabled: false
    privKeyPath: /path/to/myKey.pem
    certPath: /path/to/myCert.pem
auth:
  type: local
kms.local:
  type: local
crypto:
  keyPath: $CRIBL_HOME/local/cribl/auth/keys.json
system:
  upgrade: api
  restart: api
  installType: standalone
  intercom: true
workers:
  count: -2
  minimum: 2
  memory: 2048
proxy:
  useEnvVars: true
```

16.3.2. Breakers.Yml

Cribl's default Event Breaker Library is stored in `$CRIBL_HOME/default/cribl/breakers.yml`.

$CRIBL_HOME/default/cribl/breakers.yml

```
breaker_id: # [object]
  lib: # [string] Library
  description: # [string] Description - Brief description of this ruleset. Optional
  tags: # [string] Tags - One or more tags related to this ruleset. Optional.
  rules: # [array] Rules - List of rules. Evaluated in order, top down.
    - name: # [string] Rule Name - Rule Name.
      condition: # [string] Filter Condition - Filter expression (JS) that matches
      type: # [string] Event Breaker Type - Event Breaker Type
      timestampAnchorRegex: # [string] Timestamp Anchor - Regex to match before at
      timestamp: # [object] Timestamp Format - Auto, manual format (strptime) or c
      type: # [string] Timestamp Type
      length: # [number] Length
      format: # [string] Format
      timestampTimezone: # [string] Default timezone - Timezone to assign to timest
      timestampEarliest: # [string] Earliest timestamp allowed - The earliest times
      timestampLatest: # [string] Future timestamp allowed - The latest timestamp
      maxEventBytes: # [number] Max Event Bytes - The maximum number of bytes that
      fields: # [array] Fields - Key value pairs to be added to each event.
      - name: # [string] Name - Field Name.
        value: # [string] Value Expression - JavaScript expression to compute field
      disabled: # [boolean] Disabled - Allows breaker rule to be enabled or disable
```

16.3.3. Certificates.Yml

`certificates.yml` maintains a list of configured certificates and their parameters.

$CRIBL_HOME/local/cribl/certificates.yml

```
cerficate_id: # [object]
  description: # [string] Description - Brief description of this certificate. Opt
  cert: # [string] Certificate - Drag/drop or upload host certificate, in PEM/Base
  privKey: # [string] Private key - Certificate private key.
  passphrase: # [string] Passphrase - Passphrase. Optional.
  ca: # [string] CA certificate - Optionally, drag/drop or upload all CA certificat
  inUse: # [array of strings] Referenced - List of configurations referencing this
```

16.3.4. Groups.Yml

`groups.yml` maintains a list of groups and their configuration versions.

$CRIBL_HOME/local/cribl/groups.yml

```
group_id: # [object]
  configVersion: # [string] Config Version - Configuration version that is active f
  onPrem: # [boolean] On prem - Whether the group accepts on-prem or cloud worker r
  isFleet: # [boolean] Is Fleet - Fleet groups only manage edge nodes.
  workerRemoteAccess: # [boolean] UI access - Enable authenticated viewing of Worke
```

16.3.5. Inputs.Yml

`inputs.yml` contains settings for configuring inputs into Cribl.

$CRIBL_HOME/default/cribl/inputs.yml

```
inputs: # [object]
  collection_input: # [object]
    type: # [string] Input Type
    breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
    staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of t
    sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

    # -------------- if sendToRoutes is false ---------------

    connections: # [array] Quick Connections - Direct connections to Destinations,
      - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
        output: # [string] Destination - Select a Destination.

    # --------------------------------------------------------

    preprocess: # [object]
      disabled: # [boolean] Disabled - Enable Custom Command
      command: # [string] Command - Command to feed the data through (via stdin) a
      args: # [array of strings] Arguments - Arguments
    throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to thro
    metadata: # [array] Fields - Fields to add to events from this input.
      - name: # [string] Name - Field name
        value: # [string] Value - JavaScript expression to compute field's value, 
    disabled: # [boolean] Disabled - Enable/disable this input
    pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
    environment: # [string] Environment - Optionally, enable this config only on a
    pqEnabled: # [boolean] Enable Persistent Queue

    # -------------- if pqEnabled is true ---------------

    pq: # [object]
      mode: # [string] Mode - With Smart mode, PQ will write events to the filesys
      maxBufferSize: # [number] Max buffer size - The maximum amount of events to 
      commitFrequency: # [number] Commit frequency - The number of events to send 
      maxFileSize: # [string] Max file size - The maximum size to store in each qu
      maxSize: # [string] Max queue size - The maximum amount of disk space the qu
      path: # [string] Queue file path - The location for the persistent queue fil
      compress: # [string] Compression - Codec to use to compress the persisted da

    # --------------------------------------------------------

    streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

```
kafka_input: # [object]
  type: # [string] Input Type
  brokers: # [array of strings] Brokers - List of Kafka brokers to use, eg. local
  topics: # [array of strings] [required] Topic - Topic to subscribe to. Warning:
  groupId: # [string] Group ID - Specifies the consumer group this instance belor
  fromBeginning: # [boolean] From beginning - Whether to start reading from earl:
  kafkaSchemaRegistry: # [object] Kafka Schema Registry Authentication
    disabled: # [boolean] Disabled - Enable Schema Registry

    # -------------- if disabled is false ---------------

    schemaRegistryURL: # [string] Schema Registry URL - URL for access to the Cor
    tls: # [object] TLS settings (client side)
      disabled: # [boolean] Disabled

      # -------------- if disabled is false ---------------

      rejectUnauthorized: # [boolean] Validate server certs - Reject certs that a
      servername: # [string] Server name (SNI) - Server name for the SNI (Server
      certificateName: # [string] Certificate name - The name of the predefined c
      caPath: # [string] CA certificate path - Path on client in which to find CA
      privKeyPath: # [string] Private key path (mutual auth) - Path on client in
      certPath: # [string] Certificate path (mutual auth) - Path on client in wh:
      passphrase: # [string] Passphrase - Passphrase to use to decrypt private ke
      minVersion: # [string] Minimum TLS version - Minimum TLS version to use whe
      maxVersion: # [string] Maximum TLS version - Maximum TLS version to use whe

      # ----------------------------------------------------------


    # ----------------------------------------------------------

  connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
  requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
  sasl: # [object] Authentication - Authentication parameters to use when connect
    disabled: # [boolean] Disabled - Enable Authentication

    # -------------- if disabled is false ---------------

    mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.

    # ----------------------------------------------------------
```

```
tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ----------------------------------------------------------

sessionTimeout: # [number] Session timeout (ms) -
  Timeout used to detect client failures when using Kafka's group management fa
  If the client sends the broker no heartbeats before this timeout expires,
  the broker will remove this client from the group, and will initiate a rebala
  Value must be between the broker's configured group.min.session.timeout.ms an
  See details [here](https://kafka.apache.org/documentation/#consumerconfigs_se
rebalanceTimeout: # [number] Rebalance timeout (ms) -
  Maximum allowed time for each worker to join the group after a rebalance has
  If the timeout is exceeded, the coordinator broker will remove the worker fro
  See details [here](https://kafka.apache.org/documentation/#connectconfigs_reb
heartbeatInterval: # [number] Heartbeat interval (ms) -
  Expected time between heartbeats to the consumer coordinator when using Kafka
  Value must be lower than sessionTimeout, and typically should not exceed 1/3
  See details [here](https://kafka.apache.org/documentation/#consumerconfigs_he
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.
```

```
  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
http_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all addres
  port: # [number] [required] Port - Port to listen to.
  authTokens: # [array of strings] Auth tokens - Shared secrets to be provided by
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
```

```
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to adc
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  criblAPI: # [string] Cribl HTTP Event API - Absolute path on which to listen fc
  elasticAPI: # [string] Elasticsearch API endpoint (Bulk API) - Absolute path or
  splunkHecAPI: # [string] Splunk HEC Endpoint - Absolute path on which listen fc
  splunkHecAcks: # [boolean] Splunk HEC Acks - Whether to enable Splunk HEC ackno
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
```

```
port: # [number] [required] Port - Port to listen to.
tls: # [object] TLS settings (server side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif:
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ:
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

  # ---------------------------------------------------------

ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
authTokens: # [array] Auth tokens - Shared secrets to be provided by any Splunk
  - token: # [string] Token - Shared secrets to be provided by any Splunk forwa
    description: # [string] Description - Optional token description
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ---------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------
```

```
  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_search_input: # [object]
  searchHead: # [string] Search head - Search head base URL, can be expression, c
  search: # [string] [required] Search - Enter Splunk search here. For example:
  earliest: # [string] Earliest - The earliest time boundary for the search. Can
  latest: # [string] Latest - The latest time boundary for the search. Can be an
  cronSchedule: # [string] [required] Cron schedule - A cron schedule on which to
  endpoint: # [string] [required] Search endpoint - REST API used to create a sea
  outputMode: # [string] [required] Output mode - Format of the returned output
  endpointParams: # [array] Endpoint parameters - Optional request parameters to
    - name: # [string] Name - Parameter name
      value: # [string] Value - JavaScript expression to compute the parameter's
  endpointHeaders: # [array] Endpoint headers - Optional request headers to send
    - name: # [string] Name - Header Name
      value: # [string] Value - JavaScript expression to compute the header's val
  logLevel: # [string] Log level - Collector runtime log Level (verbosity).
  requestTimeout: # [number] Request timeout (seconds) - HTTP request inactivity
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  keepAliveTime: # [number] Keep Alive Time (seconds) - How often workers should
  maxMissedKeepAlives: # [number] Worker Timeout (periods) - The number of Keep A
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
  authType: # [string] Authentication type - Splunk Search authentication type

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication
```

```
# ------------------------------------------------------


# -------------- if authType is token ---------------

token: # [string] Token - Bearer token to include in the authorization header

# ------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Token (text secret) - Select (or create) a stored text s

# ------------------------------------------------------

type: # [string] Input Type
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
```

```
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to b
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_hec_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  authTokens: # [array] Auth tokens - Shared secrets to be provided by any client
    - token: # [string] Token - Shared secret to be provided by any client (Authc
      description: # [string] Description - Optional token description
      metadata: # [array] Fields - Fields to add to events referencing this token
      - name: # [string] Name - Field name
        value: # [string] Value - JavaScript expression to compute field's value
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
```

```
  splunkHecAPI: # [string] [required] Splunk HEC Endpoint - Absolute path on whic
  metadata: # [array] Fields - Fields to add to every event. May be overridden by
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  allowedIndexes: # [array of strings] Allowed Indexes - List values allowed in H
  splunkHecAcks: # [boolean] Splunk HEC Acks - Whether to enable Splunk HEC ackno
  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # --------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # --------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
azure_blob_input: # [object]
  type: # [string] Input Type
  queueName: # [string] Queue - The storage account queue name blob notifications
  fileFilter: # [string] Filename filter - Regex matching file names to download
  visibilityTimeout: # [number] Visibility timeout (secs) - The duration (in seco
  numReceivers: # [number] Num receivers - The Number of receiver processes to ru
  maxMessages: # [number] Max messages - The maximum number of messages to return
```

```
servicePeriodSecs: # [number] Service period (secs) - The duration (in seconds)
skipOnError: # [boolean] Skip file on error - Toggle to Yes to skip files that
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
parquetChunkSizeMB: # [number] Max Parquet chunk size (MB) - Maximum file size
parquetChunkDownloadTimeout: # [number] Parquet chunk download timeout (seconds
authType: # [string] Authentication method - Enter connection string directly,
connectionString: # [string] Connection string - Enter your Azure Storage accou

# -------------- if authType is manual ---------------



# -----------------------------------------------------

textSecret: # [string] Connection string (text secret) - Select (or create) a s

# -------------- if authType is secret ---------------



# -----------------------------------------------------

disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -----------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
```

```
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -----------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
elastic_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # -----------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  elasticAPI: # [string] [required] Elasticsearch API endpoint - Absolute path or
  authType: # [string] Authentication type - Elastic authentication type

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication
```

```
# ------------------------------------------------------

# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ------------------------------------------------------

# -------------- if authType is authTokens ---------------

authTokens: # [array of strings] Token - Bearer tokens to include in the autho

# ------------------------------------------------------

apiVersion: # [string] API Version - The API version to use for communicating v

# -------------- if apiVersion is custom ---------------

customAPIVersion: # [string] Custom API Version - Custom version information to

# ------------------------------------------------------

extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
  - name: # [string] Name - Field name
    value: # [string] Value - Field value
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
proxyMode: # [object]
  enabled: # [boolean] Enable Proxy Mode - Enable proxying of non-bulk API requ

  # -------------- if enabled is true ---------------

  url: # [string] Proxy URL - URL of the Elastic server to proxy non-bulk reque
  removeHeaders: # [array of strings] Remove headers - List of headers to remov
  timeoutSec: # [number] Proxy request timeout - Amount of time, in seconds, to
  authType: # [string] Authentication method - Enter credentials directly, or s

  # ------------------------------------------------------
```

```
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
confluent_cloud_input: # [object]
  type: # [string] Input Type
  brokers: # [array of strings] Brokers - List of Confluent Cloud brokers to use,
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
    servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
    certificateName: # [string] Certificate name - The name of the predefined cer
    caPath: # [string] CA certificate path - Path on client in which to find CA c
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
    certPath: # [string] Certificate path (mutual auth) - Path on client in which
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key
```

```
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ------------------------------------------------------

topics: # [array of strings] [required] Topic - Topic to subscribe to. Warning:
groupId: # [string] Group ID - Specifies the consumer group this instance belor
fromBeginning: # [boolean] From beginning - Whether to start reading from earl:
kafkaSchemaRegistry: # [object] Kafka Schema Registry Authentication
  disabled: # [boolean] Disabled - Enable Schema Registry

  # -------------- if disabled is false ---------------

  schemaRegistryURL: # [string] Schema Registry URL - URL for access to the Cor
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that a
    servername: # [string] Server name (SNI) - Server name for the SNI (Server
    certificateName: # [string] Certificate name - The name of the predefined c
    caPath: # [string] CA certificate path - Path on client in which to find CA
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in
    certPath: # [string] Certificate path (mutual auth) - Path on client in wh:
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private ke
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use whe
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use whe

    # ------------------------------------------------------


  # -------------------------------------------------------

connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
sasl: # [object] Authentication - Authentication parameters to use when connect
  disabled: # [boolean] Disabled - Enable Authentication

  # -------------- if disabled is false ---------------

  mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.
```

```
  # ------------------------------------------------------

sessionTimeout: # [number] Session timeout (ms) -
  Timeout used to detect client failures when using Kafka's group management fa
  If the client sends the broker no heartbeats before this timeout expires,
  the broker will remove this client from the group, and will initiate a rebala
  Value must be between the broker's configured group.min.session.timeout.ms an
  See details [here](https://kafka.apache.org/documentation/#consumerconfigs_se
rebalanceTimeout: # [number] Rebalance timeout (ms) -
  Maximum allowed time for each worker to join the group after a rebalance has
  If the timeout is exceeded, the coordinator broker will remove the worker fro
  See details [here](https://kafka.apache.org/documentation/#connectconfigs_reb
heartbeatInterval: # [number] Heartbeat interval (ms) -
  Expected time between heartbeats to the consumer coordinator when using Kafka
  Value must be lower than sessionTimeout, and typically should not exceed 1/3
  See details [here](https://kafka.apache.org/documentation/#consumerconfigs_he
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
```

```
    path: # [string] Queue file path - The location for the persistent queue fil
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
grafana_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all addres
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined ce
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certif
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept f
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept f

    # ------------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectio
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to ad
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  prometheusAPI: # [string] Remote Write API endpoint - Absolute path on which t
  lokiAPI: # [string] Logs API endpoint - Absolute path on which to listen for L
  keepAliveTimeout: # [number] Keep alive timeout (seconds) - Maximum time to wa
  prometheusAuth: # [object]
    authType: # [string] Authentication type - Remote Write authentication type

    # -------------- if authType is basic ---------------

    username: # [string] Username - Username for Basic authentication
    password: # [string] Password - Password for Basic authentication

    # ------------------------------------------------------------
```

```
    "

    # -------------- if authType is token ---------------

    token: # [string] Token - Bearer token to include in the authorization header

    # ----------------------------------------------------


    # -------------- if authType is credentialsSecret ---------------

    credentialsSecret: # [string] Credentials secret - Select (or create) a secret

    # ----------------------------------------------------


    # -------------- if authType is textSecret ---------------

    textSecret: # [string] Token (text secret) - Select (or create) a stored text

    # ----------------------------------------------------

  lokiAuth: # [object]
    authType: # [string] Authentication type - Loki logs authentication type

    # -------------- if authType is basic ---------------

    username: # [string] Username - Username for Basic authentication
    password: # [string] Password - Password for Basic authentication

    # ----------------------------------------------------


    # -------------- if authType is token ---------------

    token: # [string] Token - Bearer token to include in the authorization header

    # ----------------------------------------------------


    # -------------- if authType is credentialsSecret ---------------

    credentialsSecret: # [string] Credentials secret - Select (or create) a secret
```

```
    credentialsSecret: # [string] Credentials secret - Select (or create) a secr

    # --------------------------------------------------------


    # -------------- if authType is textSecret ---------------

    textSecret: # [string] Token (text secret) - Select (or create) a stored text

    # --------------------------------------------------------

  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # --------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # --------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
loki_input: # [object]
  type: # [string] Input Type
```

```
type: # [string] Input type
disabled: # [boolean] Disabled - Enable/disable this input
host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
port: # [number] [required] Port - Port to listen to.
tls: # [object] TLS settings (server side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept f
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept f

  # ----------------------------------------------------

maxActiveReq: # [number] Max active requests - Maximum number of active request
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectio
captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to ad
activityLogSampleRate: # [number] Activity log sample rate - How often request
requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
lokiAPI: # [string] [required] Logs API endpoint - Absolute path on which to l
authType: # [string] Authentication type - Loki logs authentication type

# -------------- if authType is basic ---------------

username: # [string] Username - Username for Basic authentication
password: # [string] Password - Password for Basic authentication

# ----------------------------------------------------


# -------------- if authType is token ---------------

token: # [string] Token - Bearer token to include in the authorization header

# ----------------------------------------------------


#                if authType is credentialSecret
```

```
# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ----------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Token (text secret) - Select (or create) a stored text s

# ----------------------------------------------------------

metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ----------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
  compress: # [string] Compression - Codec to use to compress the persisted dat

# ----------------------------------------------------------

streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

```
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
prometheus_rw_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all addres
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined ce
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certif
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept f
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept f

    # ---------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectio
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to ad
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  prometheusAPI: # [string] [required] Remote Write API endpoint - Absolute path
  keepAliveTimeout: # [number] Keep alive timeout (seconds) - Maximum time to wa
  authType: # [string] Authentication type - Remote Write authentication type

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication

  # ---------------------------------------------------------


  # -------------- if authType is token ---------------

  token: # [string] Token - Bearer token to include in the authorization header

  ..
```

```
# ------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Token (text secret) - Select (or create) a stored text s

# ------------------------------------------------------

metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send d
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
  compress: # [string] Compression - Codec to use to compress the persisted dat
```

```
  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
prometheus_input: # [object]
  dimensionList: # [array of strings] Extra Dimensions - Other dimensions to incl
  discoveryType: # [string] Discovery Type - Target discovery mechanism. Use stat

  # -------------- if discoveryType is static ---------------

  targetList: # [array of strings] Targets - List of Prometheus targets to pull r

  # ------------------------------------------------------


  # -------------- if discoveryType is dns ---------------

  nameList: # [array] DNS Names - List of DNS names to resolve
  recordType: # [string] Record Type - DNS Record type to resolve
  scrapeProtocol: # [string] Metrics Protocol - Protocol to use when collecting r
  scrapePath: # [string] Metrics Path - Path to use when collecting metrics from

  # ------------------------------------------------------


  # -------------- if discoveryType is ec2 ---------------

  usePublicIp: # [boolean] Use Public IP - Use public IP address for discovered t
  scrapeProtocol: # [string] Metrics Protocol - Protocol to use when collecting r
  scrapePort: # [number] Metrics Port - The port number in the metrics URL for di
  scrapePath: # [string] Metrics Path - Path to use when collecting metrics from
  searchFilter: # [array] Search Filter - EC2 Instance Search Filter
    - Name: # [string] Filter Name - Search filter attribute name, see: https://d
      Values: # [array of strings] Filter Values - Search Filter Values, if empty
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication
  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] Region - Region where the EC2 is located
  endpoint: # [string] Endpoint - EC2 service endpoint. If empty, defaults to AWS
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for EC2 - Use Assume Role credentials to a
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
```

```
assumeRoleExternalId: # [string] External ID - External ID to use when assuming

# ------------------------------------------------------

interval: # [number] Poll Interval - How often in minutes to scrape targets for
logLevel: # [string] [required] Log Level - Collector runtime Log Level
keepAliveTime: # [number] Keep Alive Time (seconds) - How often workers should
maxMissedKeepAlives: # [number] Worker Timeout (periods) - The number of Keep A
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
authType: # [string] Authentication method - Enter credentials directly, or sel
username: # [string] Username - Username for Prometheus Basic authentication
password: # [string] Password - Password for Prometheus Basic authentication

# -------------- if authType is manual ---------------


# ------------------------------------------------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# -------------- if authType is secret ---------------


# ------------------------------------------------------

type: # [string] Input Type
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue
```

```
  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
office365_mgmt_input: # [object]
  type: # [string] Input Type
  tenantId: # [string] Tenant ID - Office 365 Azure Tenant ID
  appId: # [string] [required] App ID - Office 365 Azure Application ID
  timeout: # [number] Timeout (secs) - HTTP request inactivity timeout, use 0 to
  keepAliveTime: # [number] Keep Alive Time (seconds) - How often workers should
  maxMissedKeepAlives: # [number] Worker Timeout (periods) - The number of Keep A
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  planType: # [string] [required] Subscription Plan - Office 365 subscription pla
  publisherIdentifier: # [string] Publisher Identifier - Optional Publisher Ident
  contentConfig: # [array] Content Types - Enable Office 365 Management Activity
    - contentType: # [string] Content Type - Office 365 Management Activity API c
      description: # [string] Interval Description - If interval type is minutes
      interval: # [number] Interval
      logLevel: # [string] Log Level - Collector runtime Log Level
      enabled: # [boolean] Enabled
  authType: # [string] Authentication method - Enter client secret directly, or s
  clientSecret: # [string] Client secret - Office 365 Azure client secret

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] Client secret (text secret) - Select (or create) a store

  # -------------- if authType is secret ---------------
```

```
  # -------------------------------------------------------

  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesys
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to
    commitFrequency: # [number] Commit frequency - The number of events to send
    maxFileSize: # [string] Max file size - The maximum size to store in each qu
    maxSize: # [string] Max queue size - The maximum amount of disk space the qu
    path: # [string] Queue file path - The location for the persistent queue fil
    compress: # [string] Compression - Codec to use to compress the persisted da

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
office365_service_input: # [object]
  type: # [string] Input Type
  tenantId: # [string] Tenant ID - Office 365 Azure Tenant ID
  appId: # [string] [required] App ID - Office 365 Azure Application ID
  timeout: # [number] Timeout (secs) - HTTP request inactivity timeout, use 0 to
  keepAliveTime: # [number] Keep Alive Time (seconds) - How often workers should
  maxMissedKeepAlives: # [number] Worker Timeout (periods) - The number of Keep
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value,
```

```
contentConfig: # [array] Content Types - Enable Office 365 Service Communicatio
  - contentType: # [string] Content Type - Office 365 Services API Content Type
    description: # [string] Interval Description - If interval type is minutes
    interval: # [number] Interval
    logLevel: # [string] Log Level - Collector runtime Log Level
    enabled: # [boolean] Enabled
authType: # [string] Authentication method - Enter client secret directly, or s
clientSecret: # [string] Client secret - Office 365 Azure client secret

# -------------- if authType is manual ---------------


# -----------------------------------------------------

textSecret: # [string] Client secret (text secret) - Select (or create) a store

# -------------- if authType is secret ---------------


# -----------------------------------------------------

disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -----------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
```

```
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
office365_msg_trace_input: # [object]
  url: # [string] Report URL - URL to use when retrieving report data.
  interval: # [number] [required] Poll interval - How often (in minutes) to run t
  startDate: # [string] Date range start - Backward offset for the search range's
  endDate: # [string] Date range end - Backward offset for the search range's tar
  logLevel: # [string] Log level - Log Level (verbosity) for collection runtime b
  timeout: # [number] Timeout (secs) - HTTP request inactivity timeout. Maximum i
  disableTimeFilter: # [boolean] Disable time filter - Disables time filtering of
  authType: # [string] Authentication method - Select authentication method.

  # -------------- if authType is manual ---------------

  username: # [string] Username - Username to run Message Trace API call.
  password: # [string] Password - Password to run Message Trace API call.

  # ------------------------------------------------------


  # -------------- if authType is secret ---------------

  credentialsSecret: # [string] Credentials secret - Select (or create) a secret

  # ------------------------------------------------------


  # -------------- if authType is oauth ---------------

  clientSecret: # [string] Client secret - client_secret to pass in the OAuth req
  tenantId: # [string] Tenant identifier - Directory ID (tenant identifier) in Mi
  clientId: # [string] Client ID - client_id to pass in the OAuth request paramet
  resource: # [string] Resource - Resource to pass in the OAuth request parameter

  # ------------------------------------------------------


  # -------------- if authType is oauthSecret ---------------
```

```
textSecret: # [string] Client secret - Select (or create) a secret that referer
tenantId: # [string] Tenant identifier - Directory ID (tenant identifier) in M
clientId: # [string] Client ID - client_id to pass in the OAuth request parame
resource: # [string] Resource - Resource to pass in the OAuth request parameter

# ------------------------------------------------------

keepAliveTime: # [number] Keep Alive Time (seconds) - How often workers should
maxMissedKeepAlives: # [number] Worker Timeout (periods) - The number of Keep A
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
type: # [string] Input Type
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
  compress: # [string] Compression - Codec to use to compress the persisted dat

# ------------------------------------------------------

streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

```
eventhub_input: # [object]
  type: # [string] Input Type
  brokers: # [array of strings] Brokers - List of Event Hubs Kafka brokers to cor
  topics: # [array of strings] [required] Event Hub name - The name of the Event
  groupId: # [string] Group ID - Specifies the consumer group this instance belor
  fromBeginning: # [boolean] From beginning - Whether to start reading from earl:
  connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
  requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
  sasl: # [object] Authentication - Authentication parameters to use when connect
    disabled: # [boolean] Disabled - Enable authentication.

    # -------------- if disabled is false ---------------

    mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.
    username: # [string] Username - The username for authentication. For Event Hu
    authType: # [string] Authentication method - Enter password directly, or sele

    # ---------------------------------------------------------

  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - For Event Hubs, this

    # ---------------------------------------------------------

  sessionTimeout: # [number] Session timeout (ms) -
    Timeout (a.k.a session.timeout.ms in Kafka domain) used to detect client fail
    If the client sends the broker no heartbeats before this timeout expires, the
    Value must be lower than rebalanceTimeout.
    See details [here](https://github.com/Azure/azure-event-hubs-for-kafka/blob/r
  rebalanceTimeout: # [number] Rebalance timeout (ms) -
    Maximum allowed time (a.k.a rebalance.timeout.ms in Kafka domain) for each wo
    If the timeout is exceeded, the coordinator broker will remove the worker fro
    See details [here](https://github.com/Azure/azure-event-hubs-for-kafka/blob/r
  heartbeatInterval: # [number] Heartbeat interval (ms) -
    Expected time (a.k.a heartbeat.interval.ms in Kafka domain) between heartbeat
    Value must be lower than sessionTimeout, and typically should not exceed 1/3
    See details [here](https://github.com/Azure/azure-event-hubs-for-kafka/blob/r
  minimizeDuplicates: # [boolean] Minimize duplicates - Enable feature to minimiz
  metadata: # [array] Fields - Fields to add to events from this input.
```

```
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send d
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
exec_input: # [object]
  disabled: # [boolean] Disabled - Enable/disable this input
  command: # [string] Command - Command to execute; supports Bourne shell syntax
  retries: # [number] Max retries - Maximum number of retry attempts in the event
  scheduleType: # [string] Schedule type - Select a schedule type; either an inte

  # -------------- if scheduleType is interval ---------------

  interval: # [number] Interval - Interval between command executions in seconds.

  # -------------------------------------------------------
```

```
  # -------------- if scheduleType is cronSchedule ---------------

  cronSchedule: # [string] Schedule - Cron schedule to execute the command on.

  # ---------------------------------------------------------

  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of t
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  type: # [string] Input Type
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesys
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
firehose_input: # [object]
  type: # [string] Input Type
```

```
disabled: # [boolean] Disabled - Enable/disable this input
host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
port: # [number] [required] Port - Port to listen to.
authTokens: # [array of strings] Auth tokens - Shared secrets to be provided by
tls: # [object] TLS settings (server side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept f
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept f

  # ----------------------------------------------------

maxActiveReq: # [number] Max active requests - Maximum number of active request
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
activityLogSampleRate: # [number] Activity log sample rate - How often request
requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -----------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------
```

```
  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ----------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
google_pubsub_input: # [object]
  type: # [string] Input Type
  topicName: # [string] Topic ID - ID of the topic to receive events from.
  subscriptionName: # [string] [required] Subscription ID - ID of the subscriptio
  createTopic: # [boolean] Create topic - If enabled, create topic if it does not
  createSubscription: # [boolean] Create subscription - If enabled, create subscr

  # -------------- if createSubscription is true ---------------

  orderedDelivery: # [boolean] Ordered delivery - If enabled, receive events in t

  # ----------------------------------------------------------

  region: # [string] Region - Region to retrieve messages from. Select 'default'
  googleAuthMethod: # [string] Authentication Method - Google authentication meth

  # -------------- if googleAuthMethod is manual ---------------

  serviceAccountCredentials: # [string] Service account credentials - Contents of

  # ----------------------------------------------------------


  # -------------- if googleAuthMethod is secret ---------------

  secret: # [string] Service account credentials (text secret) - Select (or creat

  # ----------------------------------------------------------

  maxBacklog: # [number] Max backlog - If Destination exerts backpressure, this s
```

```
  requestTimeout: # [number] Request timeout (ms) - Pull request timeout, in mil
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
cribl_input: # [object]
  type: # [string] Input Type
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------
```

```
  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
cribl_tcp_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr
```

```
    # ------------------------------------------------------

  maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
cribl_http_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  authTokens: # [array of strings] Auth tokens - Shared secrets to be provided by
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled
```

```
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif:
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ:
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

  # ---------------------------------------------------

maxActiveReq: # [number] Max active requests - Maximum number of active request
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
activityLogSampleRate: # [number] Activity log sample rate - How often request
requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ---------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
```

```
    maxFileSize: # [string] Max file size - The maximum size to store in each qu
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
tcpjson_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certif:
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ:
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
  maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  authType: # [string] Authentication method - Enter a token directly, or provide
  authToken: # [string] Auth token - Shared secret to be provided by any client (

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t
```

```
  textSecret: # [string] Auth token (text secret) - Select (or create) a stored

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesys
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to
    commitFrequency: # [number] Commit frequency - The number of events to send
    maxFileSize: # [string] Max file size - The maximum size to store in each qu
    maxSize: # [string] Max queue size - The maximum amount of disk space the qu
    path: # [string] Queue file path - The location for the persistent queue fil
    compress: # [string] Compression - Codec to use to compress the persisted da

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
system_metrics_input: # [object]
  interval: # [number] Polling interval - Time, in seconds, between consecutive
  host: # [object]
    mode: # [string]  - Select level of detail for host metrics

    # -------------- if mode is custom ---------------

    custom: # [object]
      system: # [object]
```

```
system: # [object]
  mode: # [string]  - Select the level of details for system metrics

  # -------------- if mode is custom ---------------

  processes: # [boolean] Process metrics - Generate metrics for the numbers

  # ------------------------------------------------------

cpu: # [object]
  mode: # [string]  - Select the level of details for CPU metrics

  # -------------- if mode is custom ---------------

  perCpu: # [boolean] Per CPU metrics - Generate metrics for each CPU
  detail: # [boolean] Detailed metrics - Generate metrics for all CPU state
  time: # [boolean] CPU time metrics - Generate raw, monotonic CPU time cou

  # ------------------------------------------------------

memory: # [object]
  mode: # [string]  - Select the level of details for memory metrics

  # -------------- if mode is custom ---------------

  detail: # [boolean] Detailed metrics - Generate metrics for all memory st

  # ------------------------------------------------------

network: # [object]
  mode: # [string]  - Select the level of details for network metrics

  # -------------- if mode is custom ---------------

  devices: # [array of strings] Interface filter - Network interfaces to in
  perInterface: # [boolean] Per interface metrics - Generate separate metri
  detail: # [boolean] Detailed metrics - Generate full network metrics

  # ------------------------------------------------------

disk: # [object]
  mode: # [string]  - Select the level of details for disk metrics

  #               if mode is custom
```

```
    # -------------- if mode is custom ---------------

    devices: # [array of strings] Device filter - Block devices to include/ex
    mountpoints: # [array of strings] Mountpoint filter - Filesystem mountpo
    fstypes: # [array of strings] Filesystem type filter - Filesystem types t
    perDevice: # [boolean] Per device metrics - Generate separate metrics for
    detail: # [boolean] Detailed metrics - Generate full disk metrics

    # -----------------------------------------------------


  # -------------------------------------------------------

container: # [object]
  mode: # [string]  - Select the level of detail for container metrics

  # -------------- if mode is basic ---------------

  dockerSocket: # [array of strings] Docker socket - Full paths for Docker's UN
  dockerTimeout: # [number] Docker timeout - Timeout, in seconds, for the Docke

  # -------------------------------------------------------


  # -------------- if mode is all ---------------

  dockerSocket: # [array of strings] Docker socket - Full paths for Docker's UN
  dockerTimeout: # [number] Docker timeout - Timeout, in seconds, for the Docke

  # -------------------------------------------------------


  # -------------- if mode is custom ---------------

  filters: # [array] Container Filters - Containers matching any of these will
    - expr: # [string] Expression
  allContainers: # [boolean] All containers - Include stopped and paused conta
  perDevice: # [boolean] Per device metrics - Generate separate metrics for eac
  detail: # [boolean] Detailed metrics - Generate full container metrics
  dockerSocket: # [array of strings] Docker socket - Full paths for Docker's UN
  dockerTimeout: # [number] Docker timeout - Timeout, in seconds, for the Docke

  # -------------------------------------------------------
```

```
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
persistence: # [object] persistence
  enable: # [boolean] Enable disk persistence - Persist metrics on disk

  # -------------- if enable is true ---------------

  timeWindow: # [string] Bucket time span - Time span for each file bucket
  maxDataSize: # [string] Max data size - Maximum disk space allowed to be cons
  maxDataTime: # [string] Max data age - Maximum amount of time to retain data
  compress: # [string] Compression - Select data compression format. Optional.
  destPath: # [string] Path location - Path to use to write metrics. Defaults t

  # ------------------------------------------------------

type: # [string] Input Type
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
  compress: # [string] Compression - Codec to use to compress the persisted dat
```

```
  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
system_state_input: # [object]
  interval: # [number] Polling interval - Time, in seconds, between consecutive s
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  collectors: # [object]
    hostsfile: # [object]
      enable: # [boolean] Enabled
  persistence: # [object]
    enable: # [boolean] Enable disk persistence - Persist metrics on disk

    # -------------- if enable is true ---------------

    timeWindow: # [string] Bucket time span - Time span for each file bucket
    maxDataSize: # [string] Max data size - Maximum disk space allowed to be cons
    maxDataTime: # [string] Max data age - Maximum amount of time to retain data
    compress: # [string] Compression - Select data compression format. Optional.
    destPath: # [string] Path location - Path to use to write metrics. Defaults t

    # ------------------------------------------------------

  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------
```

```
  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
windows_metrics_input: # [object]
  interval: # [number] Polling interval - Time, in seconds, between consecutive r
  host: # [object]
    mode: # [string]  - Select level of detail for host metrics

    # -------------- if mode is custom ---------------

    custom: # [object]
      system: # [object]
        mode: # [string]  - Select the level of details for system metrics

        # -------------- if mode is custom ---------------

        detail: # [boolean] Detailed metrics - Generate metrics for all system in

        # ------------------------------------------------------------

      cpu: # [object]
        mode: # [string]  - Select the level of details for CPU metrics

        # -------------- if mode is custom ---------------

        perCpu: # [boolean] Per CPU metrics - Generate metrics for each CPU
        detail: # [boolean] Detailed metrics - Generate metrics for all CPU state
        time: # [boolean] CPU time metrics - Generate raw, monotonic CPU time cou

        # ------------------------------------------------------------

      memory: # [object]
        mode: # [string]  - Select the level of details for memory metrics
```

```
      # -------------- if mode is custom ---------------

      detail: # [boolean] Detailed metrics - Generate metrics for all memory st

      # -------------------------------------------------------

    network: # [object]
      mode: # [string]  - Select the level of details for network metrics

      # -------------- if mode is custom ---------------

      devices: # [array of strings] Interface filter - Network interfaces to in
      perInterface: # [boolean] Per interface metrics - Generate separate metri
      detail: # [boolean] Detailed metrics - Generate full network metrics

      # -------------------------------------------------------

    disk: # [object]
      mode: # [string]  - Select the level of details for disk metrics

      # -------------- if mode is custom ---------------

      volumes: # [array of strings] Volume filter - Windows volumes to include/
      perVolume: # [boolean] Per volume metrics - Generate separate metrics for

      # -------------------------------------------------------


  # -----------------------------------------------------

metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
persistence: # [object] persistence
  enable: # [boolean] Enable disk persistence - Persist metrics on disk

  # -------------- if enable is true ---------------

  timeWindow: # [string] Bucket time span - Time span for each file bucket
  maxDataSize: # [string] Max data size - Maximum disk space allowed to be cons
  maxDataTime: # [string] Max data age - Maximum amount of time to retain data
  compress: # [string] Compression - Select data compression format. Optional.
  destPath: # [string] Path location - Path to use to write metrics. Defaults t
```

```
    # ------------------------------------------------------

  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
crowdstrike_input: # [object]
  type: # [string] Input Type
  queueName: # [string] Queue - The name, URL, or ARN of the SQS queue to read nc
  fileFilter: # [string] Filename filter - Regex matching file names to download
  awsAccountId: # [string] AWS Account ID - SQS queue owner's AWS account ID. Lea
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key
```

```
# ------------------------------------------------------


# -------------- if awsAuthenticationMethod is secret ---------------

awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

# -------------------------------------------------------

awsSecretKey: # [string] Secret key - Secret key
region: # [string] Region - AWS Region where the S3 bucket and SQS queue are lo
endpoint: # [string] Endpoint - S3 service endpoint. If empty, defaults to AWS
signatureVersion: # [string] Signature version - Signature version to use for s
reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
maxMessages: # [number] Max Messages - The maximum number of messages SQS shoul
visibilityTimeout: # [number] Visibility timeout seconds - The duration (in sec
numReceivers: # [number] Num receivers - The Number of receiver processes to ru
socketTimeout: # [number] Socket timeout - Socket inactivity timeout (in second
skipOnError: # [boolean] Skip file on error - Toggle to Yes to skip files that
enableAssumeRole: # [boolean] Enable for S3 - Use Assume Role credentials to ac
assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
assumeRoleExternalId: # [string] External ID - External ID to use when assuming
enableSQSAssumeRole: # [boolean] Enable for SQS - Use Assume Role credentials w
preprocess: # [object]
  disabled: # [boolean] Disabled - Enable Custom Command
  command: # [string] Command - Command to feed the data through (via stdin) ar
  args: # [array of strings] Arguments - Arguments
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pollTimeout: # [number] Poll timeout (secs) - The amount of time to wait for ev
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
```

```
    output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
datadog_agent_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
```

```
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to adc
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  extractMetrics: # [boolean] Extract metrics - Toggle to Yes to extract each inc
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  proxyMode: # [object]
    enabled: # [boolean] Forward API key validation requests - Toggle to Yes to s
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # -------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
datagen_input: # [object]
  samples: # [array] Datagen - List of datagens
    - sample: # [string] Data Generator File - Name of the datagen file
      eventsPerSec: # [number] Events Per Second Per Worker Node - Maximum no. of
  metadata: # [array] Fields - Fields to add to events from this input.
```

```
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # --------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # --------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
http_raw_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  authTokens: # [array of strings] Auth tokens - Shared secrets to be provided by
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------
```

```
  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif:
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ:
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

  # ------------------------------------------------------

maxActiveReq: # [number] Max active requests - Maximum number of active request
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
activityLogSampleRate: # [number] Activity log sample rate - How often request
requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of t:
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
allowedPaths: # [array of strings] Allowed URI paths - List of URI paths accept
allowedMethods: # [array of strings] Allowed HTTP methods - List of HTTP method
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
```

```
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
kinesis_input: # [object]
  type: # [string] Input Type
  streamName: # [string] Stream name - Kinesis stream name to read data from.
  serviceInterval: # [number] Service Period - Time interval in minutes between c
  shardExpr: # [string] Shard selection expression - A JS expression to be called
  shardIteratorType: # [string] Shard iterator start - Location at which to start
  payloadFormat: # [string] Record data format - Format of data inside the Kinesi
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] [required] Region - Region where the Kinesis stream is locat
  endpoint: # [string] Endpoint - Kinesis stream service endpoint. If empty, defa
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for Kinesis stream - Use Assume Role crede
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
  verifyKPLCheckSums: # [boolean] Verify KPL checksums - Verify Kinesis Producer
  avoidDuplicates: # [boolean] Avoid duplicate records - Yes means: when resuming
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
```

```
    value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befo
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directl

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send d
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
logstream_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all addres
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
```

```
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept f
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept f

  # ----------------------------------------------------------

ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectio
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
authType: # [string] Authentication method - Enter a token directly, or provide
authToken: # [string] Auth token - Shared secret to be provided by any client (

# -------------- if authType is manual ---------------


# ----------------------------------------------------------

textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

# -------------- if authType is secret ---------------


# ----------------------------------------------------------

pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ----------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue
```

```
  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
criblmetrics_input: # [object]
  type: # [string] Input Type
  prefix: # [string] Metric Name Prefix - A prefix that is applied to the metrics
  fullFidelity: # [boolean] Full Fidelity - Include granular metrics.  Disabling
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
```

```
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
metrics_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. For IPv4 (all addresses), use th
  udpPort: # [number] UDP Port - Enter UDP port number to listen on. Not required
  tcpPort: # [number] TCP Port - Enter TCP port number to listen on. Not required
  maxBufferSize: # [number] Max Buffer Size (events) - Maximum number of events t
  ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.
```

```
  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
s3_input: # [object]
  type: # [string] Input Type
  queueName: # [string] Queue - The name, URL, or ARN of the SQS queue to read nc
  fileFilter: # [string] Filename filter - Regex matching file names to download
  awsAccountId: # [string] AWS Account ID - SQS queue owner's AWS account ID. Lea
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ------------------------------------------------------



  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] Region - AWS Region where the S3 bucket and SQS queue are lo
  endpoint: # [string] Endpoint - S3 service endpoint. If empty, defaults to AWS
```

```
signatureVersion: # [string] Signature version - Signature version to use for s
reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
maxMessages: # [number] Max Messages - The maximum number of messages SQS shoul
visibilityTimeout: # [number] Visibility timeout seconds - The duration (in sec
numReceivers: # [number] Num receivers - The Number of receiver processes to ru
socketTimeout: # [number] Socket timeout - Socket inactivity timeout (in second
skipOnError: # [boolean] Skip file on error - Toggle to Yes to skip files that
enableAssumeRole: # [boolean] Enable for S3 - Use Assume Role credentials to ac
assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
assumeRoleExternalId: # [string] External ID - External ID to use when assuming
enableSQSAssumeRole: # [boolean] Enable for SQS - Use Assume Role credentials w
preprocess: # [object]
  disabled: # [boolean] Disabled - Enable Custom Command
  command: # [string] Command - Command to feed the data through (via stdin) ar
  args: # [array of strings] Arguments - Arguments
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
parquetChunkSizeMB: # [number] Max Parquet chunk size (MB) - Maximum file size
parquetChunkDownloadTimeout: # [number] Parquet chunk download timeout (seconds
pollTimeout: # [number] Poll timeout (secs) - The amount of time to wait for ev
disabled: # [boolean] Disabled - Enable/disable this input
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
```

```
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesys
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
snmp_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. For IPv4 (all addresses), use th
  port: # [number] [required] UDP Port - UDP port to receive SNMP traps on. Defau
  maxBufferSize: # [number] Max Buffer Size (events) - Maximum number of events t
  ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
```

```
    path: # [string] Queue file path - The location for the persistent queue fil
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
open_telemetry_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all addres
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined ce
    privKeyPath: # [string] Private key path - Path on server containing the pri
    certPath: # [string] Certificate path - Path on server containing certificat
    caPath: # [string] CA certificate path - Path on server containing CA certif
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ

    # ------------------------------------------------------

  maxActiveReq: # [number] Max active requests - Maximum number of active request
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectio
  captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to ad
  activityLogSampleRate: # [number] Activity log sample rate - How often request
  requestTimeout: # [number] Request timeout (seconds) - How long to wait for an
  extractSpans: # [boolean] Extract spans - Toggle to Yes to extract each incomi
  extractMetrics: # [boolean] Extract metrics - Toggle to Yes to extract each in
  maxActiveCxn: # [number] Max active connections - Maximum number of active con
  authType: # [string] Authentication type - OpenTelemetry authentication type

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication

  # ------------------------------------------------------


  # -------------- if authType is token ---------------
```

```
token: # [string] Token - Bearer token to include in the authorization header

# -------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# -------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Token (text secret) - Select (or create) a stored text s

# -------------------------------------------------------

metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# -------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
```

```
    maxSize: # [string] Max queue size - The maximum amount of disk space the qu
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
sqs_input: # [object]
  type: # [string] Input Type
  queueName: # [string] Queue - The name, URL, or ARN of the SQS queue to read ev
  queueType: # [string] [required] Queue Type - The queue type used (or created).

  # -------------- if queueType is standard ---------------

  numReceivers: # [number] Num receivers - The Number of receiver processes to ru

  # ------------------------------------------------------

  awsAccountId: # [string] AWS Account ID - SQS queue owner's AWS account ID. Lea
  createQueue: # [boolean] Create Queue - Create queue if it does not exist.
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] Region - AWS Region where the SQS queue is located. Required
  endpoint: # [string] Endpoint - SQS service endpoint. If empty, defaults to AWS
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for SQS - Use Assume Role credentials to a
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
```

```
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
  maxMessages: # [number] Max Messages - The maximum number of messages SQS shoul
  visibilityTimeout: # [number] Visibility Timeout Seconds - The duration (in sec
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  pollTimeout: # [number] Poll timeout (secs) - The amount of time to wait for ev
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
syslog_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. For IPv4 (all addresses), use th
  udpPort: # [number] UDP port - Enter UDP port number to listen on. Not required
  tcpPort: # [number] TCP port - Enter TCP port number to listen on. Not required
  maxBufferSize: # [number] Max buffer size (events) - Maximum number of events t
  ipWhitelistRegex: # [string] IP allowlist regex - Regex matching IP addresses t
```

```
ipWhitelistRegex: # [string] IP allowlist regex - Regex matching IP addresses t
timestampTimezone: # [string] Default timezone - Timezone to assign to timestar
singleMsgUdpPackets: # [boolean] Single msg per UDP - Whether to treat UDP pacl
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
keepFieldsList: # [array of strings] Fields to keep - Wildcard list of fields t
octetCounting: # [boolean] Octet count framing - Enable if incoming messages us
tls: # [object] TLS settings (server side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certif:
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requ:
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

  # ---------------------------------------------------------

metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false ---------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ---------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
pqEnabled: # [boolean] Enable Persistent Queue

# -------------- if pqEnabled is true ---------------

pq: # [object]
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
```

```
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
file_input: # [object]
  mode: # [string]  - Choose how to discover files to monitor.

  # -------------- if mode is manual ---------------

  path: # [string] Search path - Directory path to search for files. Environment
  depth: # [number] Max depth - Set how many subdirectories deep to search. Use 0

  # ------------------------------------------------------

  interval: # [number] Polling interval - Time, in seconds, between scanning for
  filenames: # [array of strings] Filename allowlist - The full path of discovere
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue
```

```
  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
tcp_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
  host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
  port: # [number] [required] Port - Port to listen to.
  tls: # [object] TLS settings (server side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    certificateName: # [string] Certificate name - The name of the predefined cer
    privKeyPath: # [string] Private key path - Path on server containing the priv
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    certPath: # [string] Certificate path - Path on server containing certificate
    caPath: # [string] CA certificate path - Path on server containing CA certifi
    requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
    minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

    # ------------------------------------------------------

  ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
  maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
  enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
```

```
  enableHeader: # [boolean] Enable Header - If enabled, client will pass the heac

  # -------------- if enableHeader is true ---------------

  authType: # [string] Authentication method - Enter a token directly, or provide

  # ------------------------------------------------------

  preprocess: # [object]
    disabled: # [boolean] Disabled - Enable Custom Command
    command: # [string] Command - Command to feed the data through (via stdin) ar
    args: # [array of strings] Arguments - Arguments
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send c
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
appscope_input: # [object]
  type: # [string] Input Type
  disabled: # [boolean] Disabled - Enable/disable this input
```

```
ipWhitelistRegex: # [string] IP Allowlist Regex - Regex matching IP addresses t
maxActiveCxn: # [number] Max Active Connections - Maximum number of active conn
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connectior
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
enableUnixPath: # [boolean] UNIX domain socket - Toggle to Yes to specify a fil

# -------------- if enableUnixPath is false ---------------

host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
port: # [number] Port - Port to listen to.
tls: # [object] TLS settings (server side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  certificateName: # [string] Certificate name - The name of the predefined cer
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] Certificate path - Path on server containing certificate
  caPath: # [string] CA certificate path - Path on server containing CA certifi
  requestCert: # [boolean] Authenticate client (mutual auth) - Whether to requi
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr

  # ----------------------------------------------------------


# ----------------------------------------------------------


# -------------- if enableUnixPath is true ---------------

unixSocketPath: # [string] UNIX socket path - Path to the UNIX domain socket to
unixSocketPerms: # [string,number] UNIX socket permissions - Permissions to set

# ----------------------------------------------------------

authType: # [string] Authentication method - Enter a token directly, or provide
authToken: # [string] Auth token - Shared secret to be provided by any client (
```

```
  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send d
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
wef_input: # [object]
  type: # [string] Input Type
```

```
disabled: # [boolean] Disabled - Enable/disable this input
host: # [string] Address - Address to bind on. Defaults to 0.0.0.0 (all address
port: # [number] [required] Port - Port to listen to.
tls: # [object] [required] mTLS settings
  disabled: # [boolean] Disabled - Enable TLS
  rejectUnauthorized: # [boolean] Validate client certs - Required for WEF cert
  requestCert: # [boolean] Authenticate client - Required for WEF certificate a
  certificateName: # [string] Certificate name - Name of the predefined certif
  privKeyPath: # [string] Private key path - Path on server containing the priv
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  certPath: # [string] [required] Certificate path - Path on server containing
  caPath: # [string] [required] CA certificate path - Path on server containing
  commonNameRegex: # [string] Common name - Regex matching allowable common nam
  minVersion: # [string] Minimum TLS version - Minimum TLS version to accept fr
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to accept fr
maxActiveReq: # [number] Max active requests - Maximum number of active request
enableProxyHeader: # [boolean] Enable proxy protocol - Enable if the connection
captureHeaders: # [boolean] Capture request headers - Toggle this to Yes to add
caFingerprint: # [string] CA fingerprint override - SHA1 fingerprint expected b
allowMachineIdMismatch: # [boolean] Allow MachineID mismatch - Allow events to
subscriptions: # [array] [required] Subscriptions - Subscriptions to events on
  - subscriptionName: # [string] Name - Friendly name for this subscription.
    version: # [string] Version - Version UUID for this subscription. If any su
    contentFormat: # [string] Format - Content format in which the endpoint sho
    heartbeatInterval: # [number] Heartbeat - Max time (in seconds) between end
    batchTimeout: # [number] Batch timeout - Interval (in seconds) over which t
    readExistingEvents: # [boolean] Read existing events - Set to Yes if a newl
    sendBookmarks: # [boolean] Use bookmarks - If toggled to Yes, Cribl Edge wi
    compress: # [boolean] Compression - If toggled to Yes, Stream will receive
    targets: # [array of strings] Targets - Enter the DNS names of the endpoint
    querySelector: # [string] Query builder mode - Select the query builder mod
pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

# -------------- if sendToRoutes is false --------------

connections: # [array] Quick Connections - Direct connections to Destinations,
  - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
    output: # [string] Destination - Select a Destination.

# ------------------------------------------------------

environment: # [string] Environment - Optionally, enable this config only on a
```

```
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
    mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
    maxBufferSize: # [number] Max buffer size - The maximum amount of events to h
    commitFrequency: # [number] Commit frequency - The number of events to send d
    maxFileSize: # [string] Max file size - The maximum size to store in each que
    maxSize: # [string] Max queue size - The maximum amount of disk space the que
    path: # [string] Queue file path - The location for the persistent queue file
    compress: # [string] Compression - Codec to use to compress the persisted dat

  # ---------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
win_event_logs_input: # [object]
  type: # [string] Input Type
  logNames: # [array of strings] Event Logs - Enter the event logs to collect. Ru
  readMode: # [string] Read Mode - Read all stored and future event logs, or only
  interval: # [number] Polling Interval - Time, in seconds, between checking for
  batchSize: # [number] Batch Size - Maximum number of event records to read in d
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
  disabled: # [boolean] Disabled - Enable/disable this input
  pipeline: # [string] Pipeline - Pipeline to process data from this Source befor
  sendToRoutes: # [boolean]  - Select whether to send data to Routes, or directly

  # -------------- if sendToRoutes is false ---------------

  connections: # [array] Quick Connections - Direct connections to Destinations,
    - pipeline: # [string] Pipeline/Pack - Select Pipeline or Pack. Optional.
      output: # [string] Destination - Select a Destination.

  # ---------------------------------------------------------

  environment: # [string] Environment - Optionally, enable this config only on a
  pqEnabled: # [boolean] Enable Persistent Queue

  # -------------- if pqEnabled is true ---------------

  pq: # [object]
```

```
  mode: # [string] Mode - With Smart mode, PQ will write events to the filesyst
  maxBufferSize: # [number] Max buffer size - The maximum amount of events to b
  commitFrequency: # [number] Commit frequency - The number of events to send c
  maxFileSize: # [string] Max file size - The maximum size to store in each que
  maxSize: # [string] Max queue size - The maximum amount of disk space the que
  path: # [string] Queue file path - The location for the persistent queue file
  compress: # [string] Compression - Codec to use to compress the persisted dat

# ------------------------------------------------------------

streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

16.3.6. Instance.Yml

Instance configuration is located under `$CRIBL_HOME/local/_system/instance.yml` (`C:\Program Data\Cribl\local\_system.yml` for Cribl Edge on Windows).

$CRIBL_HOME/local/_system/instance.yml or C:\Program Data\Cribl\local_system.yml for Cribl Edge on Windows

```
distributed:
  mode: # [string] One of: master | worker | single | edge | managed-edge.
  master: # [object]
    host: # [string] Instance host address.
    port: # [number] Instance post number.
    tls: # [object]
      disabled: # [boolean]
      certificateName: # [string] TLS certificate name.
      rejectUnauthorized: # [boolean] False if ignoring untrusted certificates.
      requestCert: # [boolean] Whether certificates are required and validated.
      privKeyPath: # [string] Path to private key file. Can reference $ENV_VARS.
      certPath: # [string] Path to certificate file. Can reference $ENV_VARS.
      caPath: # [string] Path to CA certificate file. Can reference $ENV_VARS.
      commonNameRegex: # [any] Allowed common names in peer certificates' subject (
      # Available values: 'TLSv1.3' | 'TLSv1.2' | 'TLSv1.1' | 'TLSv1'
      minVersion: # [string] Minimum TLS version.
      maxVersion: # [string] Maximum TLS version.
      servername: # [string] Server name.
    resiliency: # [string] Preferred failover mode, one of: none | failover.
    ipWhitelistRegex: # [string] IP addresses allowed to send data (regex).
    maxActiveCxn: # [number] Max number of active connections allowed per Worker Pr
    authToken: # [string] Token used for communication between Leader and managed r
    compression: # [string] Codec used to compress persisted data. One of: none | g
    connectionTimeout: # [number] Maximum time to wait for a connection to complete
    writeTimeout: # [number] Time (ms) to wait for a write to complete before assur
    configBundles: # [object]
      remoteUrl: # [string] Bucket to use for remote bundle storage, in s3://${buck
  group: # [string] Group the managed node belongs to.
  envRegex: # [string] Regex used to filter env variable names that can be sent by
  tags: # [array of strings] Tags associated with managed node, can be used by Lead
```

16.3.7. Jobs.Yml

`jobs.yml` maintains parameters for configured Collectors, corresponding to those listed on the UI's **Manage Collectors** page. Each collection job is listed according to the pattern shown below.

$CRIBL_HOME/local/cribl/jobs.yml

```
collection_job: # [object]
  workerAffinity: # [boolean] Worker affinity - If enabled tasks are created and ru
  collector: # [object]
    type: # [string] Collector type - The type of collector to run.

    # -------------- if type is azure_blob ---------------

    conf: # [object] [required]
      outputName: # [string] Auto-populate from - The name of the predefined Destir
      authType: # [string] Authentication method - Enter authentication data direct

      # -------------- if authType is manual ---------------

      connectionString: # [string] Connection string - Enter your Azure storage acc

      # -----------------------------------------------------


      # -------------- if authType is secret ---------------

      textSecret: # [string] Connection string (text secret) - Text secret

      # -----------------------------------------------------

      containerName: # [string] Container name - Name of the container to collect f
      path: # [string] Path - The directory from which to collect data. Templating
      extractors: # [array] Path extractors - Allows using template tokens as conte
        - key: # [string] Token - A token from the template path, e.g.: epoch
          expression: # [string] Extractor expression - JS expression that receives
      recurse: # [boolean] Recursive - Whether to recurse through subdirectories.
      maxBatchSize: # [number] Max Batch Size (objects) - Maximum number of metadat

    # -----------------------------------------------------


    # -------------- if type is filesystem ---------------

    conf: # [object] [required]
      outputName: # [string] Auto-populate from - Select a predefined configuratior
      path: # [string] Directory - The directory from which to collect data. Templa
      extractors: # [array] Path extractors - Allows using template tokens as conte
        - key: # [string] Token - A token from the template directory, e.g.: epoch
```

```
      expression: # [string] Extractor expression - JS expression that receives
  recurse: # [boolean] Recursive - Whether to recurse through subdirectories.
  maxBatchSize: # [number] Max Batch Size (files) - Maximum number of metadata

# -------------------------------------------------------


# -------------- if type is google_cloud_storage ---------------

conf: # [object] [required]
  outputName: # [string] Auto-populate from - The name of the predefined Destin
  bucket: # [string] Bucket name - Name of the bucket to collect from. This val
  path: # [string] Path - The directory from which to collect data. Templating
  extractors: # [array] Path extractors - Allows using template tokens as conte
    - key: # [string] Token - A token from the template path, e.g.: epoch
      expression: # [string] Extractor expression - JS expression that receives
  endpoint: # [string] Endpoint - Google Cloud Storage service endpoint. If emp
  disableTimeFilter: # [boolean] Disable time filter - Used to disable collecto
  recurse: # [boolean] Recursive - Whether to recurse through subdirectories.
  maxBatchSize: # [number] Max Batch Size (objects) - Maximum number of metadat
  authType: # [string] Authentication method - Enter account credentials manual

  # -------------- if authType is manual ---------------

  serviceAccountCredentials: # [string] Service account credentials - Contents

  # -------------------------------------------------------


  # -------------- if authType is secret ---------------

  textSecret: # [string] Service account credentials (text secret) - Select (or

  # -------------------------------------------------------


# -------------------------------------------------------


# -------------- if type is health_check ---------------

conf: # [object] [required]
  discovery: # [object]
```

```
  discoverType: # [string] Discover Type - Defines how task discovery will be

  # -------------- if discoverType is http ---------------

  discoverUrl: # [string] Discover URL - Expression to derive URL to use for
  discoverMethod: # [string] Discover method - Discover HTTP method.
  discoverRequestHeaders: # [array] Discover Headers - Optional discover requ
    - name: # [string] Name - Header name.
      value: # [string] Value - JavaScript expression to compute the header v
  discoverDataField: # [string] Discover Data Field - Path to field in the re

  # -------------------------------------------------------


  # -------------- if discoverType is json ---------------

  manualDiscoverResult: # [string] Discover result - Allows hard-coding the D
  discoverDataField: # [string] Discover data field - Within the response JSO

  # -------------------------------------------------------


  # -------------- if discoverType is list ---------------

  itemList: # [array of strings] Discover items - Comma-separated list of ite

  # -------------------------------------------------------

collectUrl: # [string] Health check URL - Expression to derive URL to use for
collectMethod: # [string] [required] Health check method - Health check HTTP

# -------------- if collectMethod is get ---------------

collectRequestParams: # [array] Health check parameters - Optional health che
  - name: # [string] Name - Parameter name
    value: # [string] Value - JavaScript expression to compute the parameter

# -------------------------------------------------------


# -------------- if collectMethod is post ---------------

collectRequestParams: # [array] Health check parameters - Optional health che
```

```
  - name: # [string] Name - Parameter name.
    value: # [string] Value - JavaScript expression to compute the parameter

# -------------------------------------------------------


# -------------- if collectMethod is post_with_body ---------------

collectBody: # [string] Health check POST Body - Template for POST body to se

# -------------------------------------------------------

collectRequestHeaders: # [array] Health check headers - Optional health checl
  - name: # [string] Name - Header Name
    value: # [string] Value - JavaScript expression to compute the header val
authenticateCollect: # [boolean] Authenticate health check - Enable to make a
authentication: # [string] [required] Authentication - Authentication method

# -------------- if authentication is basic ---------------

username: # [string] Username - Basic authentication username
password: # [string] Password - Basic authentication password

# -------------------------------------------------------


# -------------- if authentication is basicSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a store

# -------------------------------------------------------


# -------------- if authentication is login ---------------

loginUrl: # [string] Login URL - URL to use for login API call. This call is
username: # [string] Username - Login username
password: # [string] Password - Login password
loginBody: # [string] POST Body - Template for POST body to send with login r
tokenRespAttribute: # [string] Token Attribute - Path to token attribute in l
authHeaderExpr: # [string] Authorize Expression - JavaScript expression to co
authRequestHeaders: # [array] Authentication Headers - Optional authenticatio
  - name: # [string] Name - Header name.
```

```
    value: # [string] Value - JavaScript expression to compute the header val

# ------------------------------------------------------


# -------------- if authentication is loginSecret ---------------

loginUrl: # [string] Login URL - URL to use for login API call, this call is
credentialsSecret: # [string] Credentials secret - Select (or create) a store
loginBody: # [string] POST Body - Template for POST body to send with login r
tokenRespAttribute: # [string] Token Attribute - Path to token attribute in t
authHeaderExpr: # [string] Authorize Expression - JavaScript expression to co
authRequestHeaders: # [array] Authentication Headers - Optional authenticatio
  - name: # [string] Name - Header name.
    value: # [string] Value - JavaScript expression to compute the header val

# ------------------------------------------------------


# -------------- if authentication is oauth ---------------

loginUrl: # [string] Login URL - URL to use for the OAuth API call. This call
tokenRespAttribute: # [string] Token attribute - Path to token attribute in t
authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
clientSecretParamName: # [string] Client secret parameter - Parameter name th
clientSecretParamValue: # [string] Client secret value - Secret value to add
authRequestParams: # [array] Extra authentication parameters - OAuth request
  - name: # [string] Name - Parameter name.
    value: # [string] Value - JavaScript expression to compute the parameter
authRequestHeaders: # [array] Authentication headers - Optional authenticatio
  - name: # [string] Name - Header name.
    value: # [string] Value - JavaScript expression to compute the header's v

# ------------------------------------------------------


# -------------- if authentication is oauthSecret ---------------

loginUrl: # [string] Login URL - URL to use for the OAuth API call. This call
tokenRespAttribute: # [string] Token attribute - Path to token attribute in t
authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
clientSecretParamName: # [string] Client secret parameter - Parameter name th
textSecret: # [string] Client secret value (text secret) - Select (or create)
```

```
  authRequestParams: # [array] Extra authentication parameters - OAuth request
    - name: # [string] Name - Parameter name.
      value: # [string] Value - JavaScript expression to compute the parameter
  authRequestHeaders: # [array] Authentication headers - Optional authenticatio
    - name: # [string] Name - Header name.
      value: # [string] Value - JavaScript expression to compute the header's v

  # ------------------------------------------------------

  timeout: # [number] Request Timeout (secs) - HTTP request inactivity timeout,
  defaultBreakers: # [string] Hidden Default Breakers
  safeHeaders: # [array of strings] Safe headers - List of headers that are sa

# ------------------------------------------------------


# -------------- if type is office365_mgmt ---------------

conf: # [object] [required]
  plan_type: # [string] Subscription plan - Office 365 subscription plan for yo
  tenant_id: # [string] [required] Tenant identifier - Directory ID (tenant ide
  app_id: # [string] [required] Application identifier - Identifier of the regi
  client_secret: # [string] [required] Client secret - Application key of the r
  publisher_identifier: # [string] Publisher identifier - Optional PublisherIde
  content_type: # [string] [required] Content type - The type of content to ret

# ------------------------------------------------------


# -------------- if type is office365_service ---------------

conf: # [object] [required]
  tenant_id: # [string] Tenant identifier - Directory ID (tenant identifier) in
  app_id: # [string] [required] Application identifier - Identifier of the regi
  client_secret: # [string] [required] Client secret - Application key of the r
  content_type: # [string] [required] Content type - The type of content to ret

# ------------------------------------------------------


# -------------- if type is prometheus ---------------

conf: # [object] [required]
```

```
  dimensionList: # [array] Extra Dimensions - Other dimensions to include in ev
  username: # [string] Username - Optional username for Basic authentication
  password: # [string] Password - Optional password for Basic authentication
  discoveryType: # [string] Discovery Type - Target discovery mechanism, use st

  # -------------- if discoveryType is static ---------------

  targetList: # [array] Targets - List of Prometheus targets to pull metrics fr

  # ----------------------------------------------------------


  # -------------- if discoveryType is dns ---------------

  nameList: # [array] DNS Names - List of DNS names to resolve
  recordType: # [string] Record Type - DNS Record type to resolve
  scrapeProtocol: # [string] Metrics Protocol - Protocol to use when collecting
  scrapePath: # [string] Metrics Path - Path to use when collecting metrics fro

  # ----------------------------------------------------------


  # -------------- if discoveryType is ec2 ---------------

  usePublicIp: # [boolean] Use Public IP - Use public IP address for discovered
  scrapeProtocol: # [string] Metrics Protocol - Protocol to use when collecting
  scrapePort: # [number] Metrics Port - The port number in the metrics URL for
  scrapePath: # [string] Metrics Path - Path to use when collecting metrics fro
  searchFilter: # [array] Search Filter - EC2 Instance Search Filter
    - Name: # [string] Filter Name - Search filter attribute name, see: https:/
      Values: # [array of strings] Filter Values - Search Filter Values
  region: # [string] Region - Region from which to retrieve data.
  awsAuthenticationMethod: # [string] Authentication Method - AWS authenticatio
  enableAssumeRole: # [boolean] Enable Assume Role - Use Assume Role credential
  endpoint: # [string] Endpoint - EC2 service endpoint. If empty, defaults to A
  signatureVersion: # [string] Signature version - Signature version to use for

  # ----------------------------------------------------------


# ----------------------------------------------------------
```

```
# -------------- if type is rest ---------------

conf: # [object] [required]
  discovery: # [object]
    discoverType: # [string] Discover type - Defines how task discovery will be

    # -------------- if discoverType is http ---------------

    discoverUrl: # [string] Discover URL - Expression to derive URL to use for
    discoverMethod: # [string] Discover method - Discover HTTP method.
    discoverRequestHeaders: # [array] Discover headers - Optional discover requ
      - name: # [string] Name - Header name.
        value: # [string] Value - JavaScript expression to compute the paramete
    discoverDataField: # [string] Discover data field - Path to field in the re

    # -------------------------------------------------------


    # -------------- if discoverType is json ---------------

    manualDiscoverResult: # [string] Discover result - Allows hard-coding the D
    discoverDataField: # [string] Discover data field - Within the response JSO

    # -------------------------------------------------------


    # -------------- if discoverType is list ---------------

    itemList: # [array of strings] Discover items - Comma-separated list of ite

    # -------------------------------------------------------

  collectUrl: # [string] Collect URL - Expression to derive URL to use for the
  collectMethod: # [string] [required] Collect method - Collect HTTP method.

  # -------------- if collectMethod is get ---------------

  collectRequestParams: # [array] Collect parameters - Optional collect request
    - name: # [string] Name - Parameter name
      value: # [string] Value - JavaScript expression to compute the parameter

  # -------------------------------------------------------
```

```
# -------------- if collectMethod is post ---------------

collectRequestParams: # [array] Collect parameters - Optional collect request
  - name: # [string] Name - Parameter name.
    value: # [string] Value - JavaScript expression to compute the parameter

# -------------------------------------------------------


# -------------- if collectMethod is post_with_body ---------------

collectBody: # [string] Collect POST body - Template for POST body to send w

# -------------------------------------------------------


# -------------- if collectMethod is other ---------------

collectVerb: # [string] Collect verb - DIY HTTP verb to use for the collect
collectBody: # [string] Collect body - Template for body to send with the Co
collectRequestParams: # [array] Collect parameters - Optional collect request
  - name: # [string] Name - Parameter name.
    value: # [string] Value - JavaScript expression to compute the parameter

# -------------------------------------------------------

collectRequestHeaders: # [array] Collect headers - Optional collect request h
  - name: # [string] Name - Header Name
    value: # [string] Value - JavaScript expression to compute the header's v
pagination: # [object]
  type: # [string] Pagination - Select collect pagination scheme

  # -------------- if type is response_body ---------------

  attribute: # [array,string] Response attribute - The name of the attribute
  maxPages: # [number] Max pages - The maximum number of pages to retrieve, s

  # -------------------------------------------------------


  # -------------- if type is response_header_link ---------------
```

```
  nextRelationAttribute: # [string] Next page relation name - Relation name u
  curRelationAttribute: # [string] Current page relation name - Optional rela
  maxPages: # [number] Max pages - The maximum number of pages to retrieve, s

  # ------------------------------------------------------


  # -------------- if type is request_offset ---------------

  offsetField: # [string] Offset field name - Query string parameter that set
  offset: # [number] Starting offset - Offset index from which to start reque
  offsetSpacer: # [null]
  limitField: # [string] Limit field name - Query string parameter to set the
  limit: # [number] Limit - Maximum number of records to collect per request.
  limitSpacer: # [null]
  totalRecordField: # [string] Total record count field name - Identifies the
  maxPages: # [number] Max pages - The maximum number of pages to retrieve. S
  zeroIndexed: # [boolean] Zero-based index - Toggle to Yes to indicate that

  # ------------------------------------------------------


  # -------------- if type is request_page ---------------

  pageField: # [string] Page number field name - Query string parameter that
  page: # [number] Starting page number - Page number from which to start rec
  offsetSpacer: # [null]
  sizeField: # [string] Page size field name - Query string parameter to set
  size: # [number] Page size - Maximum number of records to collect per page.
  limitSpacer: # [null]
  totalPageField: # [string] Total page count field name - The name of the at
  totalRecordField: # [string] Total record count field name - Identifies the
  maxPages: # [number] Max pages - The maximum number of pages to retrieve. S
  zeroIndexed: # [boolean] zero-based index - Toggle to Yes to indicate that

  # ------------------------------------------------------

authentication: # [string] [required] Authentication - Authentication method

# -------------- if authentication is basic ---------------

username: # [string] Username - Basic authentication username
password: # [string] Password - Basic authentication password
```

```
# ------------------------------------------------------


# -------------- if authentication is basicSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a store

# ------------------------------------------------------


# -------------- if authentication is login ---------------

loginUrl: # [string] Login URL - URL to use for login API call. This call is
username: # [string] Username - Login username
password: # [string] Password - Login password
loginBody: # [string] POST body - Template for POST body to send with login r
tokenRespAttribute: # [string] Token attribute - Path to token attribute in T
authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
authRequestHeaders: # [array] Authentication headers - Optional authenticatio
  - name: # [string] Name - Header name.
    value: # [string] Value - JavaScript expression to compute the header's v

# ------------------------------------------------------


# -------------- if authentication is loginSecret ---------------

loginUrl: # [string] Login URL - URL to use for login API call, this call is
credentialsSecret: # [string] Credentials secret - Select (or create) a store
loginBody: # [string] POST body - Template for POST body to send with login r
tokenRespAttribute: # [string] Token attribute - Path to token attribute in T
authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
authRequestHeaders: # [array] Authentication headers - Optional authenticatio
  - name: # [string] Name - Header name.
    value: # [string] Value - JavaScript expression to compute the parameter

# ------------------------------------------------------


# -------------- if authentication is oauth ---------------

loginUrl: # [string] Login URL - URL to use for the OAuth API call. This call
```

```
  tokenRespAttribute: # [string] Token attribute - Path to token attribute in T
  authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
  clientSecretParamName: # [string] Client secret parameter - Parameter name th
  clientSecretParamValue: # [string] Client secret value - Secret value to add
  authRequestParams: # [array] Extra authentication parameters - OAuth request
    - name: # [string] Name - Parameter name.
      value: # [string] Value - JavaScript expression to compute the parameter
  authRequestHeaders: # [array] Authentication headers - Optional authenticatio
    - name: # [string] Name - Header name.
      value: # [string] Value - JavaScript expression to compute the header's v

  # -------------------------------------------------------


  # -------------- if authentication is oauthSecret ---------------

  loginUrl: # [string] Login URL - URL to use for the OAuth API call. This call
  tokenRespAttribute: # [string] Token attribute - Path to token attribute in T
  authHeaderExpr: # [string] Authorize expression - JavaScript expression to co
  clientSecretParamName: # [string] Client secret parameter - Parameter name th
  textSecret: # [string] Client secret value (text secret) - Select (or create)
  authRequestParams: # [array] Extra authentication parameters - OAuth request
    - name: # [string] Name - Parameter name.
      value: # [string] Value - JavaScript expression to compute the parameter
  authRequestHeaders: # [array] Authentication headers - Optional authenticatio
    - name: # [string] Name - Header name.
      value: # [string] Value - JavaScript expression to compute the header's v

  # -------------------------------------------------------

  timeout: # [number] Request timeout (secs) - HTTP request inactivity timeout,
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS
  disableTimeFilter: # [boolean] Disable time filter - Used to disable collecto
  safeHeaders: # [array of strings] Safe headers - List of headers that are saf

# -------------------------------------------------------


# -------------- if type is s3 ---------------

conf: # [object] [required]
  outputName: # [string] Auto-populate from - The name of the predefined Destin
  bucket: # [string] S3 bucket - S3 Bucket from which to collect data.
```

```
  region: # [string] Region - Region from which to retrieve data.
  path: # [string] Path - The directory from which to collect data. Templating
  extractors: # [array] Path extractors - Allows using template tokens as conte
    - key: # [string] Token - A token from the template path, e.g.: epoch
      expression: # [string] Extractor expression - JS expression that receives
  awsAuthenticationMethod: # [string] Authentication Method - AWS authenticatio

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key. If not present, will fall back
  awsSecretKey: # [string] Secret key - Secret key. If not present, will fall b

  # --------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret th

  # --------------------------------------------------------

  endpoint: # [string] Endpoint - S3 service endpoint. If empty, the endpoint w
  signatureVersion: # [string] Signature version - Signature version to use for
  enableAssumeRole: # [boolean] Enable Assume Role - Use Assume Role credential
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the
  assumeRoleExternalId: # [string] External ID - External ID to use when assumi
  recurse: # [boolean] Recursive - Whether to recurse through subdirectories.
  maxBatchSize: # [number] Max Batch Size (objects) - Maximum number of metadat
  reuseConnections: # [boolean] Reuse Connections - Whether to reuse connection
  rejectUnauthorized: # [boolean] Reject Unauthorized Certificates - Whether to
  verifyPermissions: # [boolean] Verify bucket permissions - Disable if you can

# --------------------------------------------------------


# -------------- if type is script ---------------

conf: # [object] [required]
  discoverScript: # [string] Discover Script - Script to discover what to colle
  collectScript: # [string] [required] Collect Script - Script to run to perfor
  shell: # [string] Shell - Shell to use to execute scripts.

# --------------------------------------------------------
```

```
# -------------- if type is splunk ---------------

conf: # [object] [required]
  searchHead: # [string] [required] Search head - Search Head base URL, can be
  search: # [string] Search - Enter Splunk search here. For example: 'index=myA
  earliest: # [string] Earliest - The earliest time boundary for the search. Ca
  latest: # [string] Latest - The latest time boundary for the search. Can be a
  endpoint: # [string] [required] Search endpoint - REST API used to create a s
  outputMode: # [string] [required] Output mode - Format of the returned output
  collectRequestParams: # [array] Extra parameters - Optional collect request p
    - name: # [string] Name - Parameter name
      value: # [string] Value - JavaScript expression to compute the parameter
  collectRequestHeaders: # [array] Extra headers - Optional collect request hea
    - name: # [string] Name - Header Name
      value: # [string] Value - JavaScript expression to compute the header's v
  authentication: # [string] [required] Authentication - Authentication method

  # -------------- if authentication is basic ---------------

  username: # [string] Username - Basic authentication username
  password: # [string] Password - Basic authentication password

  # ---------------------------------------------------------


  # -------------- if authentication is basicSecret ---------------

  credentialsSecret: # [string] Credentials secret - Select (or create) a store

  # ---------------------------------------------------------


  # -------------- if authentication is login ---------------

  loginUrl: # [string] Login URL - URL to use for login API call. This call is
  username: # [string] Username - Login username
  password: # [string] Password - Login password
  loginBody: # [string] POST Body - Template for POST body to send with login r
  tokenRespAttribute: # [string] Token Attribute - Path to token attribute in T
  authHeaderExpr: # [string] Authorize Expression - JavaScript expression to co
```

```
    # -------------------------------------------------------


    # -------------- if authentication is loginSecret ---------------

    loginUrl: # [string] Login URL - URL to use for login API call, this call is
    credentialsSecret: # [string] Credentials secret - Select (or create) a store
    loginBody: # [string] POST Body - Template for POST body to send with login r
    tokenRespAttribute: # [string] Token Attribute - Path to token attribute in l
    authHeaderExpr: # [string] Authorize Expression - JavaScript expression to co

    # -------------------------------------------------------

    timeout: # [number] Request Timeout (secs) - HTTP request inactivity timeout,
    useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS
    disableTimeFilter: # [boolean] Disable time filter - Used to disable collecto

  # -------------------------------------------------------

  destructive: # [boolean] Destructive - If set to Yes, the collector will delete
input: # [object]
  type: # [string]
  breakerRulesets: # [array of strings] Event Breaker rulesets - A list of event
  staleChannelFlushMs: # [number] Event Breaker buffer timeout - The amount of ti
  sendToRoutes: # [boolean] Send to Routes - If set to Yes, events will be sent t

  # -------------- if sendToRoutes is true ---------------

  pipeline: # [string] Pre-Processing Pipeline - Pipeline to process results befo

  # -------------------------------------------------------


  # -------------- if sendToRoutes is false ---------------

  pipeline: # [string] Pipeline - Pipeline to process results.
  output: # [string] Destination - Destination to send results to.

  # -------------------------------------------------------

  preprocess: # [object]
    disabled: # [boolean] Disabled - Enable Custom Command
    command: # [string] Command - Command to feed the data through (via stdin) an
```

```
    command: # [string] Command - Command to feed the data through (via stdin) a
    args: # [array of strings] Arguments - Arguments
  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  metadata: # [array] Fields - Fields to add to events from this input.
    - name: # [string] Name - Field name
      value: # [string] Value - JavaScript expression to compute field's value, e
type: # [string] Job type - Job type.
ttl: # [string] Time to live - Time to keep the job's artifacts on disk after job
removeFields: # [array of strings] Remove Discover fields - List of fields to rer
resumeOnBoot: # [boolean] Resume job on boot - Resumes the ad hoc job if a failu
environment: # [string] Environment - Optionally, enable this config only on a sp
schedule: # [object] Schedule - Configuration for a scheduled job.
  enabled: # [boolean] Enabled - Determines whether or not this schedule is enabl

  # -------------- if enabled is true ---------------

  cronSchedule: # [string] Cron schedule - A cron schedule on which to run this j
  maxConcurrentRuns: # [number] Max concurrent runs - The maximum number of insta
  skippable: # [boolean] Skippable - Skippable jobs can be delayed, up to their n
  run: # [object] Run settings

    # -------------- if type is collection ---------------

    rescheduleDroppedTasks: # [boolean] Reschedule tasks - Reschedule tasks that
    maxTaskReschedule: # [number] Max task reschedule - Max number of times a tas
    logLevel: # [string] Log Level - Level at which to set task logging.
    jobTimeout: # [string] Job timeout - Maximum time the job is allowed to run (
    mode: # [string] Mode - Job run mode. Preview will either return up to N matc

    # -------------- if mode is list ---------------

    discoverToRoutes: # [boolean] Send to Routes - If true, send discover results

    # -----------------------------------------------------



    # -------------- if mode is preview ---------------

    capture: # [object] Capture Settings
      duration: # [number] Capture Time (sec) - Amount of time to keep capture op
      maxEvents: # [number] Capture Up to N Events - Maximum number of events to
      level: # [string] Where to capture

    # -----------------------------------------------------
```

```
      "

      timeRangeType: # [string] Time range - Time range for scheduled job.
      earliest: # [number,string] Earliest - Earliest time, in local time.
      latest: # [number,string] Latest - Latest time, in local time.
      timestampTimezone: # [string] Range Timezone - Timezone to use for Earliest a
      expression: # [string] Filter - A filter for tokens in the provided collect p
      minTaskSize: # [string] Lower task bundle size - Limits the bundle size for s
      maxTaskSize: # [string] Upper task bundle size - Limits the bundle size for t

    # -------------------------------------------------------

  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in St
executor_job: # [object]
  executor: # [object]
    type: # [string] Executor type - The type of executor to run.
    storeTaskResults: # [boolean] Store task results - Determines whether or not to
    conf: # [object] Executor-specific settings.
  type: # [string] Job type - Job type.
  ttl: # [string] Time to live - Time to keep the job's artifacts on disk after job
  removeFields: # [array of strings] Remove Discover fields - List of fields to rem
  resumeOnBoot: # [boolean] Resume job on boot - Resumes the ad hoc job if a failur
  environment: # [string] Environment - Optionally, enable this config only on a sp
  schedule: # [object] Schedule - Configuration for a scheduled job.
    enabled: # [boolean] Enabled - Determines whether or not this schedule is enabl

    # -------------- if enabled is true ---------------

    cronSchedule: # [string] Cron schedule - A cron schedule on which to run this j
    maxConcurrentRuns: # [number] Max concurrent runs - The maximum number of insta
    skippable: # [boolean] Skippable - Skippable jobs can be delayed, up to their m
    run: # [object] Run settings

      # -------------- if type is collection ---------------

      rescheduleDroppedTasks: # [boolean] Reschedule tasks - Reschedule tasks that
      maxTaskReschedule: # [number] Max task reschedule - Max number of times a tas
      logLevel: # [string] Log Level - Level at which to set task logging.
      jobTimeout: # [string] Job timeout - Maximum time the job is allowed to run (
```

The `workerAffinity` internal parameter defaults to `false`. Cribl automatically sets this to `true` on certain jobs, specifying that collection tasks will be both created and run by the same Edge Node.

```
      discoverToRoutes: # [boolean] Send to Routes - If true, send discover results
```

```
    discoverToRoutes: # [boolean] Send to Routes - If true, send discover result

    # ------------------------------------------------------

    # -------------- if mode is preview ---------------

    capture: # [object] Capture Settings
      duration: # [number] Capture Time (sec) - Amount of time to keep capture op
      maxEvents: # [number] Capture Up to N Events - Maximum number of events to
      level: # [string] Where to capture

    # ------------------------------------------------------

    timeRangeType: # [string] Time range - Time range for scheduled job.
    earliest: # [number,string] Earliest - Earliest time, in local time.
    latest: # [number,string] Latest - Latest time, in local time.
    timestampTimezone: # [string] Range Timezone - Timezone to use for Earliest a
    expression: # [string] Filter - A filter for tokens in the provided collect p
    minTaskSize: # [string] Lower task bundle size - Limits the bundle size for s
    maxTaskSize: # [string] Upper task bundle size - Limits the bundle size for t

  # ------------------------------------------------------

streamtags: # [array of strings] Tags - Add tags for filtering and grouping in St
```

16.3.8. Job-Limits.Yml

`job-limits.yml` maintains parameters for Collection jobs and system tasks. In the UI, you can configure these per Fleet at **Fleet Settings** > **General Settings** > **Limits** > **Jobs**.

Where we do not specify or validate a minimum or maximum value, Cribl provides a prudent default. Adjust this initial value to match your architecture and needs.

$CRIBL_HOME/default/cribl/job-limits.yml

```
concurrentJobLimit: # [number; default: 10, minimum: 1] Concurrent job limit - The
concurrentSystemJobLimit: # [number; default: 10, minimum: 1] Concurrent system jol
concurrentScheduledJobLimit: # [number, default: -2] Concurrent scheduled job limi
concurrentTaskLimit: # [number; default: 2, minimum: 1] Concurrent task limit - The
concurrentSystemTaskLimit: # [number; default: 1, minimum: 1] Concurrent system tas
maxTaskPerc: # [number; default: 0.5, minimum: 0, maximum: 1] Max task usage percer
taskPollTimeoutMs: # [number; default: 60000, minimum: 10000] Task poll timeout - T
jobArtifactsReaperPeriod: # [string; default: 30 min.] Artifact reaper period - Tir
finishedJobArtifactsLimit: # [number; default: 100 items, minimum: 0] Finished job
finishedTaskArtifactsLimit: # [number; default: 500 items, minimum: 0] Finished tas
taskManifestFlushPeriodMs: # [number; default: 100, minimum: 100, maximum: 10000] M
taskManifestMaxBufferSize: # [number; default: 1000, minimum: 100, maximum: 10000]
taskManifestReadBufferSize: # [string; default: 4 KB] Manifest reader buffer size -
schedulingPolicy: # [string; default: LeastInFlight] Job dispatching - The method b
jobTimeout: # [string; default: 0] Job timeout - Maximum time a job is allowed to r
taskHeartbeatPeriod: # [number; default: 60, minimum: 60] Task heartbeat period - T
```

16.3.9. Licenses.Yml

`licenses.yml` maintains a list of Cribl Edge licenses.

$CRIBL_HOME/default/cribl/licenses.yml

```
licenses: # [array of string] - list of licenses
```

16.3.10. Limits.Yml

`limits.yml` maintains parameters for the system. In a distributed deployment, these parameters are configured on both the Leader and the Workers.

In the UI, you can configure them:

- On the Leader at **Settings** > **Global Settings** > **General Settings > Limits**.
- On a single instance at **Settings** > **General Settings > Limits**.
- On Fleets at `<Group/Fleet-name>` > **Fleet Settings** > **General Settings > Limits**.

Where we do not specify or validate a minimum or maximum value, Cribl provides a prudent default. Adjust this initial value to match your architecture and needs.

$CRIBL_HOME/default/cribl/limits.yml

```
samples: # [object] Samples
  maxSize: # [string; maximum: 3 MB; default: 256 KB] Max sample size - Maximum fi
maxMetrics: # [number; default: 1000000] Total allowed number of metric series, co
minFreeSpace: # [string; default: 5 GB] Min free disk space - The minimum amount o
maxPQSize: # [string; default: 1 TB] Maximum persistent queue size per Worker Proc
metricsGCPeriod: # [string; default: 60s] Metrics GC period - The interval on whic
metricsMaxCardinality: # [number; default: 1000] Metrics cardinality limit - The s
metricsMaxDiskSpace: # [string; default: 64GB] Metrics max disk space - Maximum al
metricsWorkerIdBlacklist: # [array of strings] Metrics worker tracking - List of m
metricsNeverDropList: # [array of strings] Metrics never-drop list - List of metri
metricsFieldsBlacklist: # [array of strings] Disable field metrics - List of event
metricsDirectory: # [string] Metrics directory - Directory in which to store metri
cpuProfileTTL: # [string; default: 30m] CPU profile TTL - The time-to-live for col
eventsMetadataSources: # [array of strings] Event metadata sources - List of event
edgeMetricsMode: # [string] Metrics to send from Edge Nodes - Choose a set of metr

# -------------- if edgeMetricsMode is custom ---------------

edgeMetricsCustomExpression: # [string] Metrics expression - JavaScript expression

# -----------------------------------------------------------

enableMetricsPersistence: # [boolean; default: true] Persist metrics - Set this to
```

16.3.11. Logger.Yml

`logger.yml` maintains logging levels and redactions, per channel. In the UI, you can configure these at **Settings** > **Global Settings** > **System** > **Logging**.

$CRIBL_HOME/default/cribl/logger.yml

```
redactFields: # [array of strings] - list of fields to redact
redactLabel: # [string] - redact label
channels: # [object] Logger channels as logger ID/log level pairs. Log levels: err
  DEFAULT: info
  input:DistMaster: debug
  output:DistWorker: debug
```

16.3.12. Mappings.Yml

Mapping ruleset configurations are stored in `$CRIBL_HOME/default/cribl/mappings.yml`.

$CRIBL_HOME/default/cribl/mappings.yml

```
mapping_ruleset_id: # [object]
  conf: # [object]
    functions: # [array] Functions - List of functions to pass data through
  active: # [boolean]
```

16.3.13. Messages.Yml

`messages.yml` stores messages displayed in the UI's `Messages` fly-out.

$CRIBL_HOME/local/cribl/logger.yml

```
message_id: # [object]
  severity: # [string] Severity
  title: # [string] Title
  text: # [string] Text
  time: # [number] Occurrence Time
  group: # [string] Group
  metadata: # [array]
```

16.3.14. Outputs.Yml

`outputs.yml` contains configuration settings for Cribl Edge Destinations.

$CRIBL_HOME/default/cribl/outputs.yml

```
outputs: # [object]
  default_output: # [object]
    type: # [string] Output Type
    defaultId: # [string,null] Default Output ID - ID of the default output. This v
    pipeline: # [string] Pipeline - Pipeline to process data before sending out to
    systemFields: # [array of strings] System fields - Set of fields to automatical
    environment: # [string] Environment - Optionally, enable this config only on a
    streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
  webhook_output: # [object]
    type: # [string] Output Type
    url: # [string] URL - URL to send events to.
    method: # [string] Method - The method to use when sending events. Defaults to
    format: # [string] Format - Specifies how to format events before sending out.

    # -------------- if format is custom ---------------

    customSourceExpression: # [string] Source expression - Expression to evaluate a
    customDropWhenNull: # [boolean] Drop when null - Whether or not to drop events
    customEventDelimiter: # [string] Event delimiter - Delimiter string to insert b
    customContentType: # [string] Content type - Content type to use for request. D

    # --------------------------------------------------------

    concurrency: # [number] Request concurrency - Maximum number of ongoing request
    maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
    maxPayloadEvents: # [number] Max events per request - Max number of events to i
    compress: # [boolean] Compress - Whether to compress the payload body before se
    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
    timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
    flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
    extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
      - name: # [string] Name - Field name
        value: # [string] Value - Field value
    useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
    failedRequestLoggingMode: # [string] Failed request logging mode - Determines v
    safeHeaders: # [array of strings] Safe headers - List of headers that are safe
    onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

    # -------------- if onBackpressure is queue ---------------

    pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
    pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
```

```
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------------

  authType: # [string] Authentication type - The authentication method to use for

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication

  # ----------------------------------------------------------


  # -------------- if authType is token ---------------

  token: # [string] Token - Bearer token to include in the authorization header

  # ----------------------------------------------------------


  # -------------- if authType is credentialsSecret ---------------

  credentialsSecret: # [string] Credentials secret - Select (or create) a secret

  # ----------------------------------------------------------


  # -------------- if authType is textSecret ---------------

  textSecret: # [string] Token (text secret) - Select (or create) a stored text s

  # ----------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
devnull_output: # [object]
  type: # [string] Output Type
```

```
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
syslog_output: # [object]
  type: # [string] Output Type
  protocol: # [string] Protocol - The network protocol to use for sending out sys

  # -------------- if protocol is tcp ---------------

  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
    servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
    certificateName: # [string] Certificate name - The name of the predefined cer
    caPath: # [string] CA certificate path - Path on client in which to find CA c
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
    certPath: # [string] Certificate path (mutual auth) - Path on client in which
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

    # ------------------------------------------------------

  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------------------------------------------------


  # -------------- if protocol is udp ---------------

  host: # [string] Address - The hostname of the receiver
  port: # [number] Port - The port to connect to on the provided host
  maxRecordSize: # [number] Max Record Size - Maximum size of syslog messages. It

  # -------------------------------------------------------
```

```
  facility: # [number] Facility - Default value for message facility, will be ove
  severity: # [number] Severity - Default value for message severity, will be ove
  appName: # [string] App Name - Default value for application name, will be ovei
  messageFormat: # [string] Message Format - The syslog message format depending
  timestampFormat: # [string] Timestamp Format - Timestamp format to use when sei
  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  octetCountFraming: # [boolean] Octet count framing - When enabled, messages wil
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_output: # [object]
  type: # [string] Output Type
  host: # [string] Address - The hostname of the receiver
  port: # [number] [required] Port - The port to connect to on the provided host
  nestedFields: # [string] Nested field serialization - Specifies how to serializ
  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
    servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
    certificateName: # [string] Certificate name - The name of the predefined cer
    caPath: # [string] CA certificate path - Path on client in which to find CA c
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
    certPath: # [string] Certificate path (mutual auth) - Path on client in which
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

    # ----------------------------------------------------

  enableMultiMetrics: # [boolean] Output multiple metrics - Output metrics in mul
  enableACK: # [boolean] Minimize in-flight data loss - Check if indexer is shutt
  maxS2Sversion: # [string] Max S2S version - The highest S2S protocol version to
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------
```

```
  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  authType: # [string] Authentication method - Enter a token directly, or provide
  authToken: # [string] Auth token - Shared secret token to use when establishing

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_lb_output: # [object]
  type: # [string] Output Type
  dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
  loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How
  maxConcurrentSenders: # [number] Max connections - Maximum number of concurrent
  nestedFields: # [string] Nested field serialization - Specifies how to serializ
  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
```

```
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ------------------------------------------------------

enableMultiMetrics: # [boolean] Output multiple metrics - Output metrics in mul
enableACK: # [boolean] Minimize in-flight data loss - Check if indexer is shutt
maxS2Sversion: # [string] Max S2S version - The highest S2S protocol version to
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ------------------------------------------------------

indexerDiscovery: # [boolean] Indexer Discovery - Automatically discover indexe

# -------------- if indexerDiscovery is true ---------------

indexerDiscoveryConfigs: # [object]  - List of configurations to set up indexer
  site: # [string] [required] Site - Clustering site of the indexers from where
  masterUri: # [string] Cluster Manager URI - Full URI of Splunk cluster Manage
  refreshIntervalSec: # [number] [required] Refresh Period - Time interval in s
  authType: # [string] Authentication method - Enter a token directly, or provi
  authToken: # [string] Auth token - Authentication token required to authentic

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------
```

```
  textSecret: # [string] Auth token (text secret) - Select (or create) a stored

  # -------------- if authType is secret ---------------


  # -------------------------------------------------------


# -----------------------------------------------------


# -------------- if indexerDiscovery is false ---------------

excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
hosts: # [array] Destinations - Set of Splunk indexers to load-balance data to.
  - host: # [string] Address - The hostname of the receiver.
    port: # [number] Port - The port to connect to on the provided host.
    tls: # [string] TLS - Whether to inherit TLS configs from group setting or
    servername: # [string] TLS Servername - Servername to use if establishing a
    weight: # [number] Load Weight - The weight to use for load-balancing purpo
dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How
maxConcurrentSenders: # [number] Max connections - Maximum number of concurrent

# -----------------------------------------------------

authType: # [string] Authentication method - Enter a token directly, or provide
authToken: # [string] Auth token - Shared secret token to use when establishing

# -------------- if authType is manual ---------------


# -------------------------------------------------------

textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

# -------------- if authType is secret ---------------


# -------------------------------------------------------

pipeline: # [string] Pipeline - Pipeline to process data before sending out to
systemFields: # [array of strings] System fields - Set of fields to automatical
```

```
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
splunk_hec_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  url: # [string] Splunk HEC Endpoint - URL to a Splunk HEC endpoint to send ever
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l

  # ------------------------------------------------------


  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
  urls: # [array] Splunk HEC Endpoints
    - url: # [string] HEC Endpoint - URL to a Splunk HEC endpoint to send events
      weight: # [number] Load Weight - The weight to use for load-balancing purpo
  dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
  loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How

  # ------------------------------------------------------


  nextQueue: # [string] Next Processing Queue - Which Splunk processing queue to
  tcpRouting: # [string] Default _TCP_ROUTING - Set the value of _TCP_ROUTING for
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  enableMultiMetrics: # [boolean] Output multi-metrics - Output metrics in multip
  authType: # [string] Authentication method - Enter a token directly, or provide

  # -------------- if authType is manual ---------------
```

```
  token: # [string] HEC Auth token - Splunk HEC authentication token

  # ------------------------------------------------------

  # -------------- if authType is secret ---------------

  textSecret: # [string] HEC Auth token (text secret) - Select (or create) a sto

  # ------------------------------------------------------

  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
tcpjson_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  host: # [string] Address - The hostname of the receiver
  port: # [number] Port - The port to connect to on the provided host

  # ------------------------------------------------------

  # -------------- if loadBalanced is true ---------------
```

```
excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
hosts: # [array] Destinations - Set of hosts to load-balance data to.
  - host: # [string] Address - The hostname of the receiver.
    port: # [number] Port - The port to connect to on the provided host.
    tls: # [string] TLS - Whether to inherit TLS configs from group setting or
    servername: # [string] TLS Servername - Servername to use if establishing a
    weight: # [number] Load Weight - The weight to use for load-balancing purpo
dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How
maxConcurrentSenders: # [number] Max connections - Maximum number of concurrent

# ------------------------------------------------------

compression: # [string] Compression - Codec to use to compress the data before
throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ------------------------------------------------------

connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
tokenTTLMinutes: # [number] Auth Token TTL minutes - The number of minutes befo
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
```

```
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # -------------------------------------------------------

  authType: # [string] Authentication method - Enter a token directly, or provide
  authToken: # [string] Auth token - Optional authentication token to include as

  # -------------- if authType is manual ---------------


  # -------------------------------------------------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

  # -------------- if authType is secret ---------------


  # -------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
wavefront_output: # [object]
  type: # [string] Output Type
  authType: # [string] Authentication method - Enter a token directly, or provide

  # -------------- if authType is manual ---------------

  token: # [string] Auth token - WaveFront API authentication token (see [here](h

  # -------------------------------------------------------


  # -------------- if authType is secret ---------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

  # -------------------------------------------------------

  domain: # [string] Domain name - WaveFront domain name, e.g. "longboard"
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
```

```
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the 1
  maxPayloadEvents: # [number] Max events per request - Max number of events to 1
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are r
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS 1
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
signalfx_output: # [object]
  type: # [string] Output Type
  authType: # [string] Authentication method - Enter a token directly, or provide

  # -------------- if authType is manual ---------------

  token: # [string] Auth token - SignalFx API access token (see [here](https://do

  # ----------------------------------------------------------


  # -------------- if authType is secret ---------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t
```

```
  # ------------------------------------------------------

  realm: # [string] Realm - SignalFx realm name, e.g. "us0"
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
filesystem_output: # [object]
  type: # [string] Output Type
  destPath: # [string] Output location - Final destination for the output files
  stagePath: # [string] Staging location - Filesystem location in which to buffer
  addIdToStagePath: # [boolean] Add output ID - Append output's ID to staging loc
  removeEmptyDirs: # [boolean] Remove staging dirs - Remove empty staging directo

  # -------------- if removeEmptyDirs is true ---------------
```

```
  emptyDirCleanupSec: # [number] Staging cleanup period - How often (secs) to cle

  # -------------------------------------------------------

  partitionExpr: # [string] Partitioning expression - JS expression defining how
  format: # [string] Data format - Format of the output data.

  # -------------- if format is json ---------------

  compress: # [string] Compress - Choose data compression format to apply before

  # -------------------------------------------------------


  # -------------- if format is parquet ---------------

  parquetSchema: # [string] Parquet schema - Select a Parquet schema. New schemas
  parquetRowGroupSize: # [string] Row group size - Ideal memory size for row grou
  parquetPageSize: # [string] Page size - Ideal memory size for page segments. E.
  spacer: # [null]
  parquetVersion: # [string] Parquet version - Determines which data types are su
  parquetDataPageVersion: # [string] Data page version - Serialization format of
  shouldLogInvalidRows: # [boolean] Log invalid rows - Output up to 20 unique rov

  # -------------------------------------------------------

  baseFileName: # [string] File name prefix expression - JavaScript expression to
  fileNameSuffix: # [string] File name suffix expression - JavaScript expression
  maxFileSizeMB: # [number] Max file size (MB) - Maximum uncompressed output file
  maxFileOpenTimeSec: # [number] Max file open time (Sec) - Maximum amount of tir
  maxFileIdleTimeSec: # [number] Max file idle time (Sec) - Maximum amount of tir
  maxOpenFiles: # [number] Max open files - Maximum number of files to keep open
  onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
s3_output: # [object]
  type: # [string] Output Type
  bucket: # [string] S3 bucket name - Name of the destination S3 bucket. Must be
  region: # [string] Region - Region where the S3 bucket is located.
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication
```

```
# -------------- if awsAuthenticationMethod is manual ---------------

awsApiKey: # [string] Access key - Access key

# -------------------------------------------------------


# -------------- if awsAuthenticationMethod is secret ---------------

awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

# -------------------------------------------------------

awsSecretKey: # [string] Secret key - Secret key
endpoint: # [string] Endpoint - S3 service endpoint. If empty, defaults to AWS
signatureVersion: # [string] Signature version - Signature version to use for s
reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
enableAssumeRole: # [boolean] Enable for S3 - Use Assume Role credentials to ac
assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
assumeRoleExternalId: # [string] External ID - External ID to use when assuming
stagePath: # [string] [required] Staging location - Filesystem location in whic
addIdToStagePath: # [boolean] Add output ID - Append output's ID to staging loc
removeEmptyDirs: # [boolean] Remove staging dirs - Remove empty staging directo

# -------------- if removeEmptyDirs is true ---------------

emptyDirCleanupSec: # [number] Staging cleanup period - How often (secs) to cle

# -------------------------------------------------------

destPath: # [string] [required] Key prefix - Prefix to append to files before u
objectACL: # [string] Object ACL - Object ACL to assign to uploaded objects.
storageClass: # [string] Storage class - Storage class to select for uploaded o
serverSideEncryption: # [string] Server side encryption - Server-side encryptio

# -------------- if serverSideEncryption is aws:kms ---------------

kmsKeyId: # [string] KMS key ID - ID or ARN of the KMS customer managed key to

# -------------------------------------------------------

partitionExpr: # [string] Partitioning expression - JS expression defining how
```

```
  partitionExpr: # [string] Partitioning expression - JS expression defining how
  format: # [string] Data format - Format of the output data.

  # -------------- if format is json ---------------

  compress: # [string] Compress - Choose data compression format to apply before

  # ------------------------------------------------------


  # -------------- if format is parquet ---------------

  parquetSchema: # [string] Parquet schema - Select a Parquet schema. New schemas
  parquetRowGroupSize: # [string] Row group size - Ideal memory size for row grou
  parquetPageSize: # [string] Page size - Ideal memory size for page segments. E.
  spacer: # [null]
  parquetVersion: # [string] Parquet version - Determines which data types are su
  parquetDataPageVersion: # [string] Data page version - Serialization format of
  shouldLogInvalidRows: # [boolean] Log invalid rows - Output up to 20 unique rov

  # ------------------------------------------------------

  baseFileName: # [string] File name prefix expression - JavaScript expression to
  fileNameSuffix: # [string] File name suffix expression - JavaScript expression
  maxFileSizeMB: # [number] Max file size (MB) - Maximum uncompressed output file
  maxFileOpenTimeSec: # [number] Max file open time (Sec) - Maximum amount of tir
  maxFileIdleTimeSec: # [number] Max file idle time (Sec) - Maximum amount of tir
  maxOpenFiles: # [number] Max open files - Maximum number of files to keep open
  onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
azure_blob_output: # [object]
  type: # [string] Output Type
  containerName: # [string] Container name - A container organizes a set of blobs
  createContainer: # [boolean] Create container - Creates the configured containe
  destPath: # [string] [required] Blob prefix - Root directory prepended to path
  stagePath: # [string] [required] Staging location - Filesystem location in whic
  addIdToStagePath: # [boolean] Add output ID - Append output's ID to staging loc
  removeEmptyDirs: # [boolean] Remove staging dirs - Remove empty staging directo

  # -------------- if removeEmptyDirs is true ---------------
```

```
emptyDirCleanupSec: # [number] Staging cleanup period - How often (secs) to cle

# -------------------------------------------------------

partitionExpr: # [string] Partitioning expression - JS expression defining how
format: # [string] Data format - Format of the output data.

# -------------- if format is json ---------------

compress: # [string] Compress - Choose data compression format to apply before

# -------------------------------------------------------


# -------------- if format is parquet ---------------

parquetSchema: # [string] Parquet schema - Select a Parquet schema. New schemas
parquetRowGroupSize: # [string] Row group size - Ideal memory size for row grou
parquetPageSize: # [string] Page size - Ideal memory size for page segments. E.
spacer: # [null]
parquetVersion: # [string] Parquet version - Determines which data types are su
parquetDataPageVersion: # [string] Data page version - Serialization format of
shouldLogInvalidRows: # [boolean] Log invalid rows - Output up to 20 unique rov

# -------------------------------------------------------

baseFileName: # [string] File name prefix expression - JavaScript expression to
fileNameSuffix: # [string] File name suffix expression - JavaScript expression
maxFileSizeMB: # [number] Max file size (MB) - Maximum uncompressed output file
maxFileOpenTimeSec: # [number] Max file open time (Sec) - Maximum amount of tir
maxFileIdleTimeSec: # [number] Max file idle time (Sec) - Maximum amount of tir
maxOpenFiles: # [number] Max open files - Maximum number of files to keep open
onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
authType: # [string] Authentication method - Enter connection string directly,
connectionString: # [string] Connection string - Enter your Azure Storage accou

# -------------- if authType is manual ---------------


# -------------------------------------------------------

textSecret: # [string] Connection string (text secret) - Select (or create) a s
```

```
  # -------------- if authType is secret ---------------


  # ----------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
azure_logs_output: # [object]
  type: # [string] Output Type
  logType: # [string] Log Type - The Log Type of events sent to this LogAnalytics
  resourceId: # [string] Resource ID - Optional Resource ID of the Azure resource
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to s
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------

  authType: # [string] Authentication method - Enter workspace ID and workspace k
  workspaceId: # [string] Workspace ID - Azure Log Analytics Workspace ID. See Az
  workspaceKey: # [string] Workspace key - Azure Log Analytics Workspace Primary
```

```
  # -------------- if authType is manual ---------------

  # -----------------------------------------------------

  keypairSecret: # [string] Secret key pair - Select (or create) a stored secret

  # -------------- if authType is secret ---------------

  # -----------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
kinesis_output: # [object]
  type: # [string] Output Type
  streamName: # [string] Stream Name - Kinesis stream name to send events to.
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # -----------------------------------------------------

  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # -----------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] [required] Region - Region where the Kinesis stream is locat
  endpoint: # [string] Endpoint - Kinesis stream service endpoint. If empty, defa
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for Kinesis stream - Use Assume Role crede
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
```

```
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
  concurrency: # [number] Put request concurrency - Maximum number of ongoing put
  maxRecordSizeKB: # [number] Max record size (KB, uncompressed) - Maximum size (
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or (

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ---------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
honeycomb_output: # [object]
  type: # [string] Output Type
  dataset: # [string] Dataset name - Name of the dataset to send events to â e.
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or (

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
```

```
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  authType: # [string] Authentication method - Enter API key directly, or select
  team: # [string] API key - Team API key where the dataset belongs

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] API key (text secret) - Select (or create) a stored text

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
azure_eventhub_output: # [object]
  type: # [string] Output Type
  brokers: # [array of strings] Brokers - List of Event Hubs Kafka brokers to cor
  topic: # [string] [required] Event Hub Name - The name of the Event Hub (a.k.a.
  ack: # [number] Acknowledgments - Control the number of required acknowledgment
  format: # [string] Record data format - Format to use to serialize events befor
  maxRecordSizeKB: # [number] Max record size (KB, uncompressed) - Maximum size (
  flushEventCount: # [number] Max events per batch - Maximum number of events in
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
  requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
  sasl: # [object] Authentication - Authentication parameters to use when connect
    disabled: # [boolean] Disabled - Enable authentication.

    # -------------- if disabled is false ---------------

    mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.
    username: # [string] Username - The username for authentication. For Event Hu
```

```
    username: # [string] Username - The username for authentication. For Event Hu
    authType: # [string] Authentication method - Enter password directly, or sele

    # ------------------------------------------------------

  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - For Event Hubs, this

    # ------------------------------------------------------

  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
google_chronicle_output: # [object]
  type: # [string] Output Type
  authenticationMethod: # [string] Authentication Method

  # -------------- if authenticationMethod is manual ---------------

  apiKey: # [string] API key - Organization's API key in Google Chronicle

  # ------------------------------------------------------


  # -------------- if authenticationMethod is secret ---------------
```

```
  apiKeySecret: # [string] API key (text secret) - Select (or create) a stored te

  # ------------------------------------------------------

  logType: # [string] Log type - Log type value to send to Chronicle. Can be over
  logTextField: # [string] Log text field - Name of the event field that contains
  logFormatType: # [string] Send events as
  region: # [string] Region - Regional endpoint to send events to
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
google_cloud_storage_output: # [object]
  type: # [string] Output Type
  bucket: # [string] Bucket name - Name of the destination Bucket. This value can
  region: # [string] [required] Region - Region where the bucket is located.
```

```
endpoint: # [string] [required] Endpoint - Google Cloud Storage service endpoi
signatureVersion: # [string] Signature version - Signature version to use for
awsAuthenticationMethod: # [string] Authentication method

# -------------- if awsAuthenticationMethod is manual ---------------

awsApiKey: # [string] Access key - HMAC access Key
awsSecretKey: # [string] Secret - HMAC secret

# --------------------------------------------------------


# -------------- if awsAuthenticationMethod is secret ---------------

awsSecret: # [string] Secret key pair - Select (or create) a stored secret tha

# --------------------------------------------------------

stagePath: # [string] [required] Staging location - Filesystem location in whi
destPath: # [string] [required] Key prefix - Prefix to append to files before
objectACL: # [string] Object ACL - Object ACL to assign to uploaded objects.
storageClass: # [string] Storage class - Storage class to select for uploaded
reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to
addIdToStagePath: # [boolean] Add output ID - Append output's ID to staging lo
removeEmptyDirs: # [boolean] Remove staging dirs - Remove empty staging direct

# -------------- if removeEmptyDirs is true ---------------

emptyDirCleanupSec: # [number] Staging cleanup period - How often (secs) to cl

# --------------------------------------------------------

partitionExpr: # [string] Partitioning expression - JS expression defining how
format: # [string] Data format - Format of the output data.

# -------------- if format is json ---------------

compress: # [string] Compress - Choose data compression format to apply before

# --------------------------------------------------------
```

```
  # -------------- if format is parquet ---------------

  parquetSchema: # [string] Parquet schema - Select a Parquet schema. New schemas
  parquetRowGroupSize: # [string] Row group size - Ideal memory size for row grou
  parquetPageSize: # [string] Page size - Ideal memory size for page segments. E.
  spacer: # [null]
  parquetVersion: # [string] Parquet version - Determines which data types are su
  parquetDataPageVersion: # [string] Data page version - Serialization format of
  shouldLogInvalidRows: # [boolean] Log invalid rows - Output up to 20 unique rov

  # -----------------------------------------------------

  baseFileName: # [string] File name prefix expression - JavaScript expression to
  fileNameSuffix: # [string] File name suffix expression - JavaScript expression
  maxFileSizeMB: # [number] Max file size (MB) - Maximum uncompressed output file
  maxFileOpenTimeSec: # [number] Max file open time (Sec) - Maximum amount of tir
  maxFileIdleTimeSec: # [number] Max file idle time (Sec) - Maximum amount of tir
  maxOpenFiles: # [number] Max open files - Maximum number of files to keep open
  onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
google_pubsub_output: # [object]
  type: # [string] Output Type
  topicName: # [string] Topic ID - ID of the topic to send events to.
  createTopic: # [boolean] Create topic - If enabled, create topic if it does not
  orderedDelivery: # [boolean] Ordered delivery - If enabled, send events in the
  region: # [string] Region - Region to publish messages to. Select 'default' to
  googleAuthMethod: # [string] Authentication Method - Google authentication meth

  # -------------- if googleAuthMethod is manual ---------------

  serviceAccountCredentials: # [string] Service account credentials - Contents of

  # -----------------------------------------------------



  # -------------- if googleAuthMethod is secret ---------------

  secret: # [string] Service account credentials (text secret) - Select (or creat

  # -----------------------------------------------------
```

```
  batchSize: # [number] Batch size - The maximum number of items the Google API s
  batchTimeout: # [number] Batch timeout (ms) - The maximum amount of time, in mi
  maxQueueSize: # [number] Max queue size - Maximum number of queued batches befo
  maxRecordSizeKB: # [number] Max batch size (KB) - Maximum size (KB) of batches
  maxInProgress: # [number] Max concurrent requests - The maximum number of in-pr
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
kafka_output: # [object]
  type: # [string] Output Type
  brokers: # [array of strings] Brokers - List of Kafka brokers to connect to, e.
  topic: # [string] [required] Topic - The topic to publish events to. Can be ove
  ack: # [number] Acknowledgments - Control the number of required acknowledgment
  format: # [string] Record data format - Format to use to serialize events befor
  compression: # [string] Compression - Codec to use to compress the data before
  maxRecordSizeKB: # [number] Max record size (KB, uncompressed) - Maximum size (
  flushEventCount: # [number] Max events per batch - Maximum number of events in
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  kafkaSchemaRegistry: # [object] Kafka Schema Registry Authentication
    disabled: # [boolean] Disabled - Enable Schema Registry

    # -------------- if disabled is false ---------------

    schemaRegistryURL: # [string] Schema Registry URL - URL for access to the Cor
    defaultKeySchemaId: # [number] Default key schema ID - Used when __keySchemaI
    defaultValueSchemaId: # [number] Default value schema ID - Used when __valueS
    tls: # [object] TLS settings (client side)
      disabled: # [boolean] Disabled
```

```
    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that a
    servername: # [string] Server name (SNI) - Server name for the SNI (Server
    certificateName: # [string] Certificate name - The name of the predefined c
    caPath: # [string] CA certificate path - Path on client in which to find CA
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in
    certPath: # [string] Certificate path (mutual auth) - Path on client in wh
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private ke
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use whe
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use whe

    # ------------------------------------------------------


  # ------------------------------------------------------

connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
sasl: # [object] Authentication - Authentication parameters to use when connect
  disabled: # [boolean] Disabled - Enable Authentication

  # -------------- if disabled is false ---------------

  mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.

  # ------------------------------------------------------

tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when
```

```
    # ------------------------------------------------------

  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # -------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
confluent_cloud_output: # [object]
  type: # [string] Output Type
  brokers: # [array of strings] Brokers - List of Confluent Cloud brokers to conn
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
    servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
    certificateName: # [string] Certificate name - The name of the predefined cer
    caPath: # [string] CA certificate path - Path on client in which to find CA c
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
    certPath: # [string] Certificate path (mutual auth) - Path on client in which
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

    # -------------------------------------------------------

  topic: # [string] [required] Topic - The topic to publish events to. Can be ove
  ack: # [number] Acknowledgments - Control the number of required acknowledgment
  format: # [string] Record data format - Format to use to serialize events befor
```

```
compression: # [string] Compression - Codec to use to compress the data before
maxRecordSizeKB: # [number] Max record size (KB, uncompressed) - Maximum size (
flushEventCount: # [number] Max events per batch - Maximum number of events in
flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
kafkaSchemaRegistry: # [object] Kafka Schema Registry Authentication
  disabled: # [boolean] Disabled - Enable Schema Registry

  # -------------- if disabled is false ---------------

  schemaRegistryURL: # [string] Schema Registry URL - URL for access to the Cor
  defaultKeySchemaId: # [number] Default key schema ID - Used when __keySchemaI
  defaultValueSchemaId: # [number] Default value schema ID - Used when __valueS
  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that a
    servername: # [string] Server name (SNI) - Server name for the SNI (Server
    certificateName: # [string] Certificate name - The name of the predefined c
    caPath: # [string] CA certificate path - Path on client in which to find CA
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in
    certPath: # [string] Certificate path (mutual auth) - Path on client in whi
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private ke
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use whe
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use whe

    # ----------------------------------------------------------


  # ----------------------------------------------------------

connectionTimeout: # [number] Connection timeout (ms) - Maximum time to wait fo
requestTimeout: # [number] Request timeout (ms) - Maximum time to wait for a su
sasl: # [object] Authentication - Authentication parameters to use when connect
  disabled: # [boolean] Disabled - Enable Authentication

  # -------------- if disabled is false ---------------

  mechanism: # [string] SASL mechanism - SASL authentication mechanism to use.

  # ----------------------------------------------------------
```

```
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ---------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
elastic_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  url: # [string] Bulk API URL or Cloud ID - Enter Cloud ID or URL to an Elastic
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l

  # ---------------------------------------------------------



  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
  urls: # [array] Bulk API URLs
    - url: # [string] URL - URL to an Elastic node to send events to â e.g., ht
      weight: # [number] Load Weight - The weight to use for load-balancing purpo
  dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
  loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How

  # ---------------------------------------------------------

  index: # [string] Index or Data Stream - Index or Data Stream to send events to
  docType: # [string] Type - Document type to use for events. Can be overwritten
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
```

```
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the 1
  maxPayloadEvents: # [number] Max events per request - Max number of events to :
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are r
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  extraParams: # [array] Extra Parameters - Extra Parameters.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  auth: # [object]
    disabled: # [boolean] Authentication Disabled

    # -------------- if disabled is false ---------------

    authType: # [string] Authentication method - Enter credentials directly, or s

    # ---------------------------------------------------

  elasticVersion: # [string] Elastic Version - Optional Elasticsearch version, us
  elasticPipeline: # [string] Elastic pipeline - Optional Elasticsearch destinati
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ---------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
newrelic_output: # [object]
```

```
type: # [string] Output Type
region: # [string] Region - Which New Relic region endpoint to use.
logType: # [string] Log type - Name of the logtype to send with events, e.g.: c
messageField: # [string] Log message field - Name of field to send as log messa
metadata: # [array] Fields - Fields to add to events from this input.
  - name: # [string] Name - Field name
    value: # [string] Value - JavaScript expression to compute field's value, e
concurrency: # [number] Request concurrency - Maximum number of ongoing request
maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the p
maxPayloadEvents: # [number] Max events per request - Max number of events to i
compress: # [boolean] Compress - Whether to compress the payload body before se
rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
  - name: # [string] Name - Field name
    value: # [string] Value - Field value
useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
safeHeaders: # [array of strings] Safe headers - List of headers that are safe
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ----------------------------------------------------------

authType: # [string] Authentication method - Enter API key directly, or select
apiKey: # [string] API key - New Relic API key. Can be overridden using __newRe

# -------------- if authType is manual ---------------



# ----------------------------------------------------------

textSecret: # [string] API key (text secret) - Select (or create) a stored text
```

```
  # -------------- if authType is secret ---------------


  # -------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
newrelic_events_output: # [object]
  type: # [string] Output Type
  region: # [string] [required] Region - Which New Relic region endpoint to use.
  accountId: # [string] Account ID - New Relic account ID
  eventType: # [string] [required] Event type - Default eventType to use when not
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # -------------------------------------------------------

  authType: # [string] Authentication method - Enter API key directly, or select
  apiKey: # [string] API key - New Relic API key. Can be overridden using __newRe
```

```
  # -------------- if authType is manual ---------------


  # -----------------------------------------------------

  textSecret: # [string] API key (text secret) - Select (or create) a stored text

  # -------------- if authType is secret ---------------


  # -----------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
influxdb_output: # [object]
  type: # [string] Output Type
  url: # [string] Write API URL - URL of an InfluxDB cluster to send events to, e
  useV2API: # [boolean] Use v2 API - The v2 API can be enabled with InfluxDB vers

  # -------------- if useV2API is false ---------------

  database: # [string] Database - Database to write to.

  # -----------------------------------------------------


  # -------------- if useV2API is true ---------------

  bucket: # [string] Bucket - Bucket to write to.
  org: # [string] Organization - Organization ID for this bucket.

  # -----------------------------------------------------

  timestampPrecision: # [string] Timestamp precision - Sets the precision for the
  dynamicValueFieldName: # [boolean] Dynamic value fields - Enabling this will pu
  valueFieldName: # [string] Value field name - Name of the field in which to sto
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the p
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
```

```
timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
  - name: # [string] Name - Field name
    value: # [string] Value - Field value
useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
safeHeaders: # [array of strings] Safe headers - List of headers that are safe
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ---------------------------------------------------------

authType: # [string] Authentication type - InfluxDB authentication type

# -------------- if authType is basic ---------------

username: # [string] Username - Username for Basic authentication
password: # [string] Password - Password for Basic authentication

# ---------------------------------------------------------


# -------------- if authType is token ---------------

token: # [string] Token - Bearer token to include in the authorization header

# ---------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ---------------------------------------------------------
```

```
  # -------------- if authType is textSecret ---------------

  textSecret: # [string] Token (text secret) - Select (or create) a stored text s

  # ---------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
cloudwatch_output: # [object]
  type: # [string] Output Type
  logGroupName: # [string] Log group name - CloudWatch log group to associate eve
  logStreamName: # [string] [required] Log stream prefix - Prefix for CloudWatch
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ---------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ---------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] [required] Region - Region where the CloudWatchLogs is locat
  endpoint: # [string] Endpoint - CloudWatchLogs service endpoint. If empty, defa
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for CloudWatchLogs - Use Assume Role crede
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
  maxQueueSize: # [number] Max queue size - Maximum number of queued batches befo
  maxRecordSizeKB: # [number] Max record size (KB, uncompressed) - Maximum size (
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
```

```
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
minio_output: # [object]
  type: # [string] Output Type
  endpoint: # [string] [required] MinIO endpoint - MinIO service url (e.g. http:/
  bucket: # [string] MinIO bucket name - Name of the destination MinIO bucket. Th
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] Region - Region where the MinIO service/cluster is located
  stagePath: # [string] [required] Staging location - Filesystem location in whic
  addIdToStagePath: # [boolean] Add output ID - Append output's ID to staging loc
  removeEmptyDirs: # [boolean] Remove staging dirs - Remove empty staging directo

  # -------------- if removeEmptyDirs is true ---------------
```

```
#                    # removeEmptyDirs is true

emptyDirCleanupSec: # [number] Staging cleanup period - How often (secs) to cle

# ------------------------------------------------------

destPath: # [string] [required] Key prefix - Root directory to prepend to path
signatureVersion: # [string] Signature version - Signature version to use for s
objectACL: # [string] Object ACL - Object ACL to assign to uploaded objects.
storageClass: # [string] Storage class - Storage class to select for uploaded o
serverSideEncryption: # [string] Server-side encryption - Server-side encryptio
reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
partitionExpr: # [string] Partitioning expression - JS expression defining how
format: # [string] Data format - Format of the output data.

# -------------- if format is json ---------------

compress: # [string] Compress - Choose data compression format to apply before

# ------------------------------------------------------


# -------------- if format is parquet ---------------

parquetSchema: # [string] Parquet schema - Select a Parquet schema. New schemas
parquetRowGroupSize: # [string] Row group size - Ideal memory size for row grou
parquetPageSize: # [string] Page size - Ideal memory size for page segments. E.
spacer: # [null]
parquetVersion: # [string] Parquet version - Determines which data types are su
parquetDataPageVersion: # [string] Data page version - Serialization format of
shouldLogInvalidRows: # [boolean] Log invalid rows - Output up to 20 unique rov

# ------------------------------------------------------

baseFileName: # [string] File name prefix expression - JavaScript expression to
fileNameSuffix: # [string] File name suffix expression - JavaScript expression
maxFileSizeMB: # [number] Max file size (MB) - Maximum uncompressed output file
maxFileOpenTimeSec: # [number] Max file open time (Sec) - Maximum amount of tir
maxFileIdleTimeSec: # [number] Max file idle time (Sec) - Maximum amount of tir
maxOpenFiles: # [number] Max open files - Maximum number of files to keep open
onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
pipeline: # [string] Pipeline - Pipeline to process data before sending out to
systemFields: # [array of strings] System fields - Set of fields to automatical
```

```
  systemFields: # [array of strings] System fields - Set of fields to automatica
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
statsd_output: # [object]
  type: # [string] Output Type
  protocol: # [string] Destination Protocol - Protocol to use when communicating

  # -------------- if protocol is tcp ---------------

  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # ------------------------------------------------------

  host: # [string] [required] Host - The hostname of the destination.
  port: # [number] [required] Port - Destination port.
  mtu: # [number] Max record Size (Bytes) - Used when Protocol is UDP, to specify
  flushPeriodSec: # [number] Flush period (sec) - Used when Protocol is TCP, to s
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
statsd_ext_output: # [object]
  type: # [string] Output Type
  protocol: # [string] Destination Protocol - Protocol to use when communicating

  # -------------- if protocol is tcp ---------------

  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # ------------------------------------------------------

  host: # [string] [required] Host - The hostname of the destination.
  port: # [number] [required] Port - Destination port.
  mtu: # [number] Max record Size (Bytes) - Used when Protocol is UDP, to specify
  flushPeriodSec: # [number] Flush period (sec) - Used when Protocol is TCP, to s
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
```

```
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
graphite_output: # [object]
  type: # [string] Output Type
  protocol: # [string] Destination Protocol - Protocol to use when communicating

  # -------------- if protocol is tcp ---------------

  throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
  connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
  writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------------------------------------------------

  host: # [string] [required] Host - The hostname of the destination.
  port: # [number] [required] Port - Destination port.
  mtu: # [number] Max record Size (Bytes) - Used when Protocol is UDP, to specify
  flushPeriodSec: # [number] Flush period (sec) - Used when Protocol is TCP, to s
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
router_output: # [object]
  type: # [string] Output Type
  rules: # [array] Rules - Event routing rules
    - filter: # [string] Filter Expression - JavaScript expression to select even
      output: # [string] Output - Output to send matching events to
      description: # [string] Description - Description of this rule's purpose
      final: # [boolean] Final - Flag to control whether to stop the event from b
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
sqs_output: # [object]
  type: # [string] Output Type
  queueName: # [string] Queue Name - The name, URL, or ARN of the SQS queue to se
  queueType: # [string] [required] Queue Type - The queue type used (or created).
  awsAccountId: # [string] AWS Account ID - SQS queue owner's AWS account ID. Lea
  messageGroupId: # [string] Message Group ID - This parameter applies only to FI
  createQueue: # [boolean] Create Queue - Create queue if it does not exist.
  awsAuthenticationMethod: # [string] Authentication method - AWS authentication

  # -------------- if awsAuthenticationMethod is manual ---------------
```

```
  # -------------- if awsAuthenticationMethod is manual ---------------

  awsApiKey: # [string] Access key - Access key

  # ------------------------------------------------------


  # -------------- if awsAuthenticationMethod is secret ---------------

  awsSecret: # [string] Secret key pair - Select (or create) a stored secret that

  # ------------------------------------------------------

  awsSecretKey: # [string] Secret key - Secret key
  region: # [string] Region - AWS Region where the SQS queue is located. Required
  endpoint: # [string] Endpoint - SQS service endpoint. If empty, defaults to AWS
  signatureVersion: # [string] Signature version - Signature version to use for s
  reuseConnections: # [boolean] Reuse connections - Whether to reuse connections
  rejectUnauthorized: # [boolean] Reject unauthorized certificates - Whether to r
  enableAssumeRole: # [boolean] Enable for SQS - Use Assume Role credentials to a
  assumeRoleArn: # [string] AssumeRole ARN - Amazon Resource Name (ARN) of the ro
  assumeRoleExternalId: # [string] External ID - External ID to use when assuming
  maxQueueSize: # [number] Max queue size - Maximum number of queued batches befo
  maxRecordSizeKB: # [number] Max record size (KB) - Maximum size (KB) of batches
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  maxInProgress: # [number] Max concurrent requests - The maximum number of in-pr
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
snmp_output: # [object]
```

```
snmp_output: # [object]
  type: # [string] Output Type
  hosts: # [array] SNMP Trap Destinations - One or more SNMP destinations to forw
    - host: # [string] Address - Destination host
      port: # [number] Port - Destination port, default is 162
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
sumo_logic_output: # [object]
  type: # [string] Output Type
  url: # [string] API URL - Sumo Logic HTTP collector URL to which events should
  customSource: # [string] Custom source name - Optionally, override the source n
  customCategory: # [string] Custom source category - Optionally, override the so
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the n
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ---------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

```
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
datadog_output: # [object]
  type: # [string] Output Type
  contentType: # [string] Send logs as - The content type to use when sending log
  message: # [string] Message field - Name of the event field that contains the m
  source: # [string] Source - Name of the source to send with logs. When you send
  host: # [string] Host - Name of the host to send with logs. When you send logs
  service: # [string] Service - Name of the service to send with logs. When you s
  tags: # [array of strings] Datadog tags - List of tags to send with logs (e.g.,
  allowApiKeyFromEvents: # [boolean] Allow API key from events - If enabled, the
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
  severity: # [string] Severity - Default value for message severity. When you se
  site: # [string] Datadog site
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the p
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ---------------------------------------------------------

  authType: # [string] Authentication method - Enter API key directly, or select
  apiKey: # [string] API key - Organization's API key in Datadog

  # -------------- if authType is manual ---------------
```

```
  # ------------------------------------------------------

  textSecret: # [string] API key (text secret) - Select (or create) a stored text

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
grafana_cloud_output: # [object]
  type: # [string] Output Type
  lokiUrl: # [string] Loki URL - The endpoint to send logs to, e.g.: https://logs
  prometheusUrl: # [string] [required] Prometheus URL - The remote_write endpoint
  message: # [string] Logs message field - Name of the event field that contains
  messageFormat: # [string] Message Format - Which format to use when sending log

  # -------------- if messageFormat is json ---------------

  compress: # [boolean] Compress - Whether to compress the payload body before se

  # ------------------------------------------------------

  labels: # [array] Logs labels - List of labels to send with logs. Labels define
    - name: # [string] Name - Name of the label.
      value: # [string] Value - Value of the label.
  metricRenameExpr: # [string] Metrics renaming expression - A JS expression that
  prometheusAuth: # [object]
    authType: # [string] Authentication Type - The authentication method to use t

    # -------------- if authType is token ---------------

    token: # [string] Auth token - Bearer token to include in the authorization h

    # ------------------------------------------------------


    # -------------- if authType is textSecret ---------------

    # [string] Auth token (text secret) - Select (or create) a stored text secret
```

```
    textSecret: # [string] Auth token (text secret) - Select (or create) a stored
    # ------------------------------------------------------

    # -------------- if authType is basic ---------------

    username: # [string] Username - Username for authentication
    password: # [string] Password - Password (a.k.a API key in Grafana Cloud doma

    # ------------------------------------------------------

    # -------------- if authType is credentialsSecret ---------------

    credentialsSecret: # [string] Credentials secret - Select (or create) a secre

    # ------------------------------------------------------

  lokiAuth: # [object]
    authType: # [string] Authentication Type - The authentication method to use t

    # -------------- if authType is token ---------------

    token: # [string] Auth token - Bearer token to include in the authorization h

    # ------------------------------------------------------

    # -------------- if authType is textSecret ---------------

    textSecret: # [string] Auth token (text secret) - Select (or create) a stored

    # ------------------------------------------------------

    # -------------- if authType is basic ---------------

    username: # [string] Username - Username for authentication
    password: # [string] Password - Password (a.k.a API key in Grafana Cloud doma

    # ------------------------------------------------------
```

```
    # -------------- if authType is credentialsSecret ---------------

    credentialsSecret: # [string] Credentials secret - Select (or create) a secre

    # ----------------------------------------------------

  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the i
  maxPayloadEvents: # [number] Max events per request - Maximum number of events
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  systemFields: # [array of strings] System fields - Set of fields to automatical
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
loki_output: # [object]
  type: # [string] Output Type
  url: # [string] Loki URL - The endpoint to send logs to.
  message: # [string] Logs message field - Name of the event field that contains
  messageFormat: # [string] Message Format - Which format to use when sending log

  # -------------- if messageFormat is json ---------------
```

```
compress: # [boolean] Compress - Whether to compress the payload body before se

# -------------------------------------------------------

labels: # [array] Logs labels - List of labels to send with logs. Labels define
  - name: # [string] Name - Name of the label.
    value: # [string] Value - Value of the label.
authType: # [string] Authentication Type - The authentication method to use for

# -------------- if authType is token ---------------

token: # [string] Auth token - Bearer token to include in the authorization hea

# -------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

# -------------------------------------------------------


# -------------- if authType is basic ---------------

username: # [string] Username - Username for authentication
password: # [string] Password - Password (a.k.a API key in Grafana Cloud domain

# -------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# -------------------------------------------------------

concurrency: # [number] Request concurrency - Maximum number of ongoing request
maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the p
maxPayloadEvents: # [number] Max events per request - Maximum number of events
rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
```

```
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  systemFields: # [array of strings] System fields - Set of fields to automatical
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
prometheus_output: # [object]
  type: # [string] Output Type
  url: # [string] Remote Write URL - The endpoint to send metrics to.
  metricRenameExpr: # [string] Metric renaming expression - A JS expression that
  sendMetadata: # [boolean] Send metadata - Whether to generate and send metadata

  # -------------- if sendMetadata is true ---------------

  metricsFlushPeriodSec: # [number] Metadata flush period (sec) - How frequently

  # ----------------------------------------------------------

  systemFields: # [array of strings] System fields - Set of fields to automatical
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
```

```
extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
  - name: # [string] Name - Field name
    value: # [string] Value - Field value
useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS 
failedRequestLoggingMode: # [string] Failed request logging mode - Determines 
safeHeaders: # [array of strings] Safe headers - List of headers that are safe
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or 

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ----------------------------------------------------------

authType: # [string] Authentication type - Remote Write authentication type

# -------------- if authType is basic ---------------

username: # [string] Username - Username for Basic authentication
password: # [string] Password - Password for Basic authentication

# ----------------------------------------------------------


# -------------- if authType is token ---------------

token: # [string] Token - Bearer token to include in the authorization header

# ----------------------------------------------------------


# -------------- if authType is credentialsSecret ---------------

credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# ----------------------------------------------------------
```

```
  # -------------- if authType is textSecret ---------------

  textSecret: # [string] Token (text secret) - Select (or create) a stored text s

  # ---------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
ring_output: # [object]
  type: # [string] Output Type
  format: # [string] Data format - Format of the output data.
  partitionExpr: # [string] Partitioning expression - JS expression to define hov
  maxDataSize: # [string] Max data size - Maximum disk space allowed to be consur
  maxDataTime: # [string] Max data age - Maximum amount of time to retain data (e
  compress: # [string] Compression - Select data compression format. Optional.
  destPath: # [string] Path location - Path to use to write metrics. Defaults to
  onBackpressure: # [string] Backpressure behavior - Whether to block or drop eve
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
open_telemetry_output: # [object]
  type: # [string] Output Type
  endpoint: # [string] Endpoint - The endpoint to send OTEL events to. It can be
  authType: # [string] Authentication type - OpenTelemetry authentication type

  # -------------- if authType is basic ---------------

  username: # [string] Username - Username for Basic authentication
  password: # [string] Password - Password for Basic authentication

  # ---------------------------------------------------------


  # -------------- if authType is token ---------------

  token: # [string] Token - Bearer token to include in the authorization header

  # ---------------------------------------------------------


  # -------------- if authType is credentialsSecret ---------------
```

```
credentialsSecret: # [string] Credentials secret - Select (or create) a secret

# -------------------------------------------------------


# -------------- if authType is textSecret ---------------

textSecret: # [string] Token (text secret) - Select (or create) a stored text s

# -------------------------------------------------------

tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # -------------------------------------------------------

metadata: # [array] Metadata - Extra information to send with each gRPC request
  - key: # [string] Key - The key of the metadata.
    value: # [string] Value - The value of the metadata.
concurrency: # [number] Request concurrency - Maximum number of ongoing request
maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
safeHeaders: # [array of strings] Safe headers - List of headers that are safe
connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
keepAliveTime: # [number] Keep Alive Time (seconds) - How often the sender shou
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

# -------------- if onBackpressure is queue ---------------
```

```
  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
dataset_output: # [object]
  type: # [string] Output Type
  messageField: # [string] Message field - Name of the event field that contains
  excludeFields: # [array of strings] Exclude fields - Fields to exclude from the
  serverHostField: # [string] Server/host field - Name of the event field that co
  timestampField: # [string] Timestamp field - Name of the event field that conta
  defaultSeverity: # [string] Severity - Default value for event severity. If the
  site: # [string] DataSet site - DataSet site to which events should be sent
  customUrl: # [string]
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the i
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  compress: # [boolean] Compress - Whether to compress the payload body before se
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
```

```
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ------------------------------------------------------

  authType: # [string] Authentication method - Enter API key directly, or select
  apiKey: # [string] API key - A 'Log Write Access' API key for the DataSet accou

  # -------------- if authType is manual ---------------


  # ------------------------------------------------------

  textSecret: # [string] API key (text secret) - Select (or create) a stored text

  # -------------- if authType is secret ---------------


  # ------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
logstream_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  host: # [string] Address - The hostname of the receiver
  port: # [number] Port - The port to connect to on the provided host

  # ------------------------------------------------------


  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
  hosts: # [array] Destinations - Set of hosts to load-balance data to.
    - host: # [string] Address - The hostname of the receiver.
      port: # [number] Port - The port to connect to on the provided host.
      tls: # [string] TLS - Whether to inherit TLS configs from group setting or
```

```
    servername: # [string] TLS Servername - Servername to use if establishing a
    weight: # [number] Load Weight - The weight to use for load-balancing purpo
dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How
maxConcurrentSenders: # [number] Max connections - Maximum number of concurrent

# ------------------------------------------------------

compression: # [string] Compression - Codec to use to compress the data before
throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ------------------------------------------------------

connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
tokenTTLMinutes: # [number] Auth Token TTL minutes - The number of minutes befo
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ------------------------------------------------------
```

```
  authType: # [string] Authentication method - Enter a token directly, or provide
  authToken: # [string] Auth token - Optional authentication token to include as

  # -------------- if authType is manual ---------------


  # -----------------------------------------------------

  textSecret: # [string] Auth token (text secret) - Select (or create) a stored t

  # -------------- if authType is secret ---------------


  # -----------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
cribl_tcp_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  host: # [string] Address - The hostname of the receiver
  port: # [number] Port - The port to connect to on the provided host

  # -----------------------------------------------------


  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
  hosts: # [array] Destinations - Set of hosts to load-balance data to.
    - host: # [string] Address - The hostname of the receiver.
      port: # [number] Port - The port to connect to on the provided host.
      tls: # [string] TLS - Whether to inherit TLS configs from group setting or
      servername: # [string] TLS Servername - Servername to use if establishing a
      weight: # [number] Load Weight - The weight to use for load-balancing purpo
  dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
  loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How
  maxConcurrentSenders: # [number] Max connections - Maximum number of concurrent
```

```
# ----------------------------------------------------------

compression: # [string] Compression - Codec to use to compress the data before
throttleRatePerSec: # [string] Throttling - Rate (in bytes per second) to throt
tls: # [object] TLS settings (client side)
  disabled: # [boolean] Disabled

  # -------------- if disabled is false ---------------

  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
  servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
  certificateName: # [string] Certificate name - The name of the predefined cer
  caPath: # [string] CA certificate path - Path on client in which to find CA c
  privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
  certPath: # [string] Certificate path (mutual auth) - Path on client in which
  passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
  minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
  maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

  # ----------------------------------------------------------

connectionTimeout: # [number] Connection Timeout - Amount of time (milliseconds
writeTimeout: # [number] Write Timeout - Amount of time (milliseconds) to wait
tokenTTLMinutes: # [number] Auth Token TTL minutes - The number of minutes befo
excludeFields: # [array of strings] Exclude Fields - Fields to exclude from the
onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or c

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# ----------------------------------------------------------

pipeline: # [string] Pipeline - Pipeline to process data before sending out to
systemFields: # [array of strings] System fields - Set of fields to automatical
environment: # [string] Environment - Optionally, enable this config only on a
streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

```
cribl_http_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  url: # [string] Cribl endpoint - URL of a Cribl Worker to send events to, e.g.,
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l

  # ------------------------------------------------------



  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
  urls: # [array] Cribl Worker Endpoints
    - url: # [string] Cribl Endpoint - URL of a Cribl Worker to send events to, e
      weight: # [number] Load Weight - The weight to use for load-balancing purpo
  dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
  loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - How

  # ------------------------------------------------------

  tls: # [object] TLS settings (client side)
    disabled: # [boolean] Disabled

    # -------------- if disabled is false ---------------

    rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are
    servername: # [string] Server name (SNI) - Server name for the SNI (Server Na
    certificateName: # [string] Certificate name - The name of the predefined cer
    caPath: # [string] CA certificate path - Path on client in which to find CA c
    privKeyPath: # [string] Private key path (mutual auth) - Path on client in wh
    certPath: # [string] Certificate path (mutual auth) - Path on client in which
    passphrase: # [string] Passphrase - Passphrase to use to decrypt private key.
    minVersion: # [string] Minimum TLS version - Minimum TLS version to use when
    maxVersion: # [string] Maximum TLS version - Maximum TLS version to use when

    # ------------------------------------------------------

  tokenTTLMinutes: # [number] Auth Token TTL minutes - The number of minutes befo
  excludeFields: # [array of strings] Exclude fields - Fields to exclude from the
  compression: # [string] Compression - Codec to use to compress the data before
```

```
  concurrency: # [number] Request concurrency - Maximum number of ongoing request
  maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
  maxPayloadEvents: # [number] Max events per request - Max number of events to i
  rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
  timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
  flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
  extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
    - name: # [string] Name - Field name
      value: # [string] Value - Field value
  failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
  safeHeaders: # [array of strings] Safe headers - List of headers that are safe
  streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
  onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

  # -------------- if onBackpressure is queue ---------------

  pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
  pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
  pqPath: # [string] Queue file path - The location for the persistent queue file
  pqCompress: # [string] Compression - Codec to use to compress the persisted dat
  pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
  pqControls: # [object]

  # ----------------------------------------------------------

  pipeline: # [string] Pipeline - Pipeline to process data before sending out to
  systemFields: # [array of strings] System fields - Set of fields to automatical
  environment: # [string] Environment - Optionally, enable this config only on a
humio_hec_output: # [object]
  type: # [string] Output Type
  loadBalanced: # [boolean] Load balancing - Use load-balanced destinations

  # -------------- if loadBalanced is false ---------------

  url: # [string] Humio HEC Endpoint - URL to a Humio HEC endpoint to send events
  useRoundRobinDns: # [boolean] Round-robin DNS - Enable to use round-robin DNS l

  # ----------------------------------------------------------


  # -------------- if loadBalanced is true ---------------

  excludeSelf: # [boolean] Exclude current host IPs - Exclude all IPs of the curr
```

```
urls: # [array] Humio HEC Endpoints
  - url: # [string] HEC Endpoint - URL to a Humio HEC endpoint to send events t
    weight: # [number] Load Weight - The weight to use for load-balancing purpo
dnsResolvePeriodSec: # [number] DNS resolution period (seconds) - Re-resolve ar
loadBalanceStatsPeriodSec: # [number] Load balance stats period (seconds) - Hov

# -------------------------------------------------------

concurrency: # [number] Request concurrency - Maximum number of ongoing request
maxPayloadSizeKB: # [number] Max body size (KB) - Maximum size, in KB, of the r
maxPayloadEvents: # [number] Max events per request - Max number of events to i
compress: # [boolean] Compress - Whether to compress the payload body before se
rejectUnauthorized: # [boolean] Validate server certs - Reject certs that are n
timeoutSec: # [number] Request timeout - Amount of time, in seconds, to wait fo
flushPeriodSec: # [number] Flush period (sec) - Maximum time between requests.
extraHttpHeaders: # [array] Extra HTTP headers - Extra HTTP headers.
  - name: # [string] Name - Field name
    value: # [string] Value - Field value
failedRequestLoggingMode: # [string] Failed request logging mode - Determines w
safeHeaders: # [array of strings] Safe headers - List of headers that are safe
format: # [string] Request Format - Send data in JSON format to the api/v1/inge
authType: # [string] Authentication method - Enter a token directly, or provide

# -------------- if authType is manual ---------------

token: # [string] HEC Auth token - Humio HEC authentication token

# -------------------------------------------------------


# -------------- if authType is secret ---------------

textSecret: # [string] HEC Auth token (text secret) - Select (or create) a stor

# -------------------------------------------------------

onBackpressure: # [string] Backpressure behavior - Whether to block, drop, or q

# -------------- if onBackpressure is queue ---------------

pqMaxFileSize: # [string] Max file size - The maximum size to store in each que
pqMaxSize: # [string] Max queue size - The maximum amount of disk space the que
pqPath: # [string] Queue file path - The location for the persistent queue file
```

```
pqCompress: # [string] Compression - Codec to use to compress the persisted dat
pqOnBackpressure: # [string] Queue-full behavior - Whether to block or drop eve
pqControls: # [object]

# -------------------------------------------------------

pipeline: # [string] Pipeline - Pipeline to process data before sending out to
systemFields: # [array of strings] System fields - Set of fields to automatical
environment: # [string] Environment - Optionally, enable this config only on a
streamtags: # [array of strings] Tags - Add tags for filtering and grouping in
```

16.3.15. Parsers.Yml

`parsers.yml` stores configuration data for the Knowledge > Parsers Library.

$CRIBL_HOME/default/cribl/parsers.yml

```
parser_id: # [object]
  lib: # [string] Library
  description: # [string] Description - Brief description of this parser. Optional
  tags: # [string] Tags - One or more tags related to this parser. Optional.
  type: # [string] Type - Parser/Formatter type to use.
  fields: # [array of strings] List of Fields - Fields expected to be extracted, ir
```

16.3.16. Perms.Yml

`perms.yml` stores Permissions configured for Cribl Edge users.

```
perms:
  - gid:    # [string]
    id:     # [string] Not used if type is 'groups'.
    type:   # [string] Resource type, one of: 'groups' | 'datasets' | 'dataset-prov
    policy: # [string] Policy name.
```

16.3.17. Policies.Yml

`policies.yml` contains RBAC Policy definitions. Default example:

$CRIBL_HOME/default/cribl/policies.yml

```
GroupFull:
  args:
    - groupName
  template:
    - PATCH /master/groups/${groupName}/deploy
    - GroupEdit ${groupName}
GroupEdit:
  args:
    - groupName
  template:
    - '* /m/${groupName}'
    - '* /m/${groupName}/*'
    - GroupRead ${groupName}
GroupCollect:
  args:
    - groupName
  template:
    - POST /m/${groupName}/lib/jobs
    - PATCH /m/${groupName}/lib/jobs/*
    - POST /m/${groupName}/jobs
    - PATCH /m/${groupName}/jobs/*
    - GroupRead ${groupName}
GroupRead:
  args:
    - groupName
  template:
    - GET /m/${groupName}
    - GET /m/${groupName}/*
    - POST /m/${groupName}/preview
    - POST /m/${groupName}/system/capture
    - POST /m/${groupName}/lib/expression
    - GET /master/groups/${groupName}
    - GET /master/workers
    - GET /master/workers/*
    - '* /w/*'
    - GET /master/groups
    - GET /system/info
    - GET /system/info/*
    - GET /system/logs
    - GET /system/logs/group/${groupName}/*
    - GET /system/settings
    - GET /system/settings/*
```

- GET /system/instance/distributed
- GET /system/instance/distributed/*
- GET /version
- GET /version/*
- GET /version/info
- GET /version/info/*
- GET /version/status
- GET /version/status/*
- GET /mappings
- GET /mappings/*
- GET /system/messages
- GET /system/message/*
- GET /ui/*
- POST /system/metrics/query
- GET /clui
- POST /system/capture

16.3.18. Regexes.Yml

`regexes.yml` maintains a list of regexes. Cribl's Regex Library ships under `default`. Each regex is listed according to the following pattern:

$CRIBL_HOME/default/cribl/regexes.yml

```
egex_id: # [object]
  lib: # [string] Library
  description: # [string] Description - Brief description of this regex. Optional.
  regex: # [string] Regex pattern - Regex pattern. Required.
  sampleData: # [string] Sample data - Sample data for this regex. Optional.
  tags: # [string] Tags - One or more tags related to this regex. Optional.
```

16.3.19. Roles.Yml

`roles.yml` contains RBAC Role definitions. Default example:

$CRIBL_HOME/default/cribl/roles.yml

```
admin:
  description: 'Members with admin role have permission to do anything and everyth
  policy:
    - '* *'
reader_all:
  description: 'Members with reader_all role get read-only access to all Worker Gr
  policy:
    - GroupRead *
collect_all:
  description: 'Members of this group can run existing collection jobs of all Work
  policy:
    - GroupCollect *
editor_all:
  description: 'Members with editor_all role get read/write access to all Worker G
  policy:
    - GroupEdit *
owner_all:
  description: 'Members with owner_all role get read/write access as well as Deplo
  policy:
    - GroupFull *
user:
  description: 'The base user role allows users to see the system info along with
  policy:
    - GET /system/info
    - GET /system/info/*
    - GET /system/users
    - GET /system/instance/distributed
    - GET /system/instance/distributed/*
    - GET /clui
    - PATCH /ui/*
```

16.3.20. Samples.Yml

`samples.yml` contains metadata about about stored sample data files (size, number of events, date created, name, etc.). Each sample is listed according to the following pattern:

$CRIBL_HOME/local/cribl/samples.yml

```
sample_id: # [object]
  sampleName: # [string] File Name - Filename to save the sample as. Required.
  pipelineId: # [string] Associate with Pipeline - Select a pipeline to associate w
  description: # [string] Description - Brief description of this sample file. Opti
  ttl: # [number] Expiration (hours) - Time to live for the sample, the TTL is rese
  tags: # [string] Tags - One or more tags related to this sample file. Optional.
```

The corresponding sample files reside in `$CRIBL_HOME/data/samples`.

16.3.21. Schemas.Yml

`schemas.yml` stores configuration data for the Knowledge > Schema Library.

$CRIBL_HOME/default/cribl/schemas.yml

```
schema_id: # [object]
  description: # [string] Description - Brief description of this schema. Optional
  schema: # [string] Schema - JSON schema matching standards of draft version 2019-
```

16.3.22. Scripts.Yml

`scripts.yml` stores configuration data for scripts configured at **Settings** > **Global Settings** > **Scripts**:

$CRIBL_HOME/local/cribl/scripts.yml

```
script_id: # [object]
  command: # [string] Command - Command to execute for this script
  description: # [string] Description - Brief description of this script. Optional
  args: # [array of strings] Arguments - Arguments to pass when executing this scr
  env: # [object] Env Variables - Extra environment variables to set when executing
```

16.3.23. Secrets.Yml

`secrets.yml` stores secrets for Cribl Edge.

The secrets are decrypted and encrypted using a `cribl.secret` file. `cribl.secret` is unique per Fleet and resides in the `$CRIBL_HOME/groups/<group-name>/local/cribl/auth` directory on Leader nodes, or in `$CRIBL_HOME/local/cribl/auth` for Workers or single-instance deployments.

```
<secret-id>:
  secretType: # [string] One of: keypair | text | credentials
  secrets:
    # -------------- if secretType is keypair ---------------
    apiKey: "<hashed-value>"
    secretKey: "<hashed-value>"

    # -------------- if secretType is text ------------------
    value: "<hashed-value>"

    # -------------- if secretType is credentials -----------
    username: "<hashed-value>"
    password: "<hashed-value>"
```

16.3.24. Service.Yml

`service.yml` maintains configuration for Cribl Edge service processes.

In the UI, you can configure them on the Leader at **Settings** > **Global Settings** > **System** > **Service Processes** > **Services**.

Be careful about reducing the predefined limits. Extremely low values can prevent some Cribl Edge components from functioning.

$CRIBL_HOME/default/cribl/service.yml

```
connections: # [Object] - Heap memory limit for the connection listener processes.
  procs: # [number; default: 1, minimum: 1]
  memoryLimit: # [string; default: 2GB]
metrics: # [Object] - Heap memory limit for the metrics process
  procs: # [number; default: 1] — Single process; values other than 1 will be treat
  memoryLimit: # [string; default: 2GB]
notifications: # [Object] - Heap memory limit for the notifications process
  procs: # [number; default: 1] — Single process; values other than 1 will be treat
  memoryLimit: # [string; default: 2GB]
```

16.3.25. VARS.YML

`vars.yml` stores configuration data for the Knowledge > Global Variables Library.

$CRIBL_HOME/default/cribl/vars.yml

```
variable_id: # [object]
  lib: # [string] Library
  description: # [string] Description - Brief description of this variable. Optiona
  type: # [string] Type - Type of variable.
  value: # [string] Value - Value of variable
  tags: # [string] Tags - One or more tags related to this variable. Optional.
```

16.4. Cribl Expressions

Native Cribl Edge methods can be found under `C.*`, and can be invoked from any Function that allows for expression evaluations. For example, to create a field that is the SHA1 of a another field's value, you can use the Eval Function with this **Evaluate Fields** pair:

Name	Value Expression
myNewField	`C.Mask.sha1(myOtherField)`

Where fields' names contain special characters, you can reference them using the `__e['<field-name-here>']` convention. For details, see Fields with Non-Alphanumeric Characters.

Cribl expressions offer methods and properties in the following classes:

- C.Crypto - data encryption and decryption
- C.Encode and C.Decode - encoding and decoding data
- C.Lookup - running lookups
- C.Mask - masking sensitive data
- C.Net - network methods
- C.Text - text manipulation
- C.Time - time and date manipulation
- C.Schema - schema validation
- C.Secret - secret management
- C.env - environment properties
- C.os - system methods
- C.vars - global variables
- C.version - Cribl Edge versions
- C.Misc - various other methods

16.4.1. C.Crypto – Data Encryption And Decryption

C.Crypto.decrypt()

```
Crypto.decrypt(value: string, escape: boolean = false, escapeSeq = '""'): string
```

Decrypts all occurrences of ciphers in the given value. Instances that cannot be decrypted (for any reason) are left intact.

Returns `value` with ciphers decrypted.

Parameter	Type	Description
`value`	string	String in which to look for ciphers.
`escape`	boolean	Set to `true` to escape double quotes in output after decryption (for example, for data encrypted in Splunk.) Defaults to `false`.
`escapeSeq`	string	String used to escape double quotes. The default `'""'` escapes CSV output.

Examples

To decrypt the contents of the `encrypted` field, use:

```
C.Crypto.decrypt(encrypted)
```

C.Crypto.encrypt()

```
Crypto.encrypt(value: Buffer | string, keyclass: number, keyId?: string, defaultVa
```

Encrypts the given value with the `keyId`, or with a `keyId` picked up automatically based on `keyclass`.

Returns the encrypted value. If encryption does not succeed, returns `defaultVal` if specified; otherwise, `value`.

Parameter	Type	Description
`value`	string | Buffer	The value to encrypt.
`keyclass`	number	Key Class to pick a key from, if `keyId` isn't specified.
`keyId`	string	Encryption [`keyId``](securing-data-encryption#encryption-keys), takes precedence over keyclass`.
`defaultVal`	string	Value to return if encryption fails; if unspecified, the original value is returned.

Examples

To encrypt the `customer` field using key with `keyId` 2, use:

```
C.Crypto.encrypt(customer, -1, 2)
```

To encrypt the same field with a key selected from the Key Class 3, without specifying a `keyId`:

```
C.Crypto.encrypt(customer, 3)
```

C.Crypto.createHmac()

```
Crypto.createHmac(value: string | Buffer, secret: string, algorithm: string = 'sha2
```

Generates an HMAC that can be added to events, or can be used to validate events that contain an HMAC.

Returns the calculated HMAC digest on success; otherwise, `value`.

Parameter	Type	Description
`value`	string | Buffer	The data to encrypt. When the `outputFormat` is invalid or undefined, this parameter is returned as the digest, via a Buffer.
`secret`	string	The secret key used to generate the MAC.
`algorithm`	string	The hash algorithm used to generate the MAC. Defaults to `'sha256'`. Run `openssl list -digest-algorithms` to see

Parameter	Type	Description
		the list of available algorithms.
`outputFormat`	`'base64' | 'hex' | 'latin1'`	Output format for the MAC. Defaults to `'hex'`.

Examples

To generate HMAC from the `customer` field, provide the secret as the second argument:

```
C.Crypto.createHmac(customer, 'kETvYtLKZkgIzBvfRdVcZULBITANjOMd')
```

To additionally specify that you want to use the SHA512 hash algorithm and to output the HMAC in `latin1` format, run:

```
C.Crypto.createHmac(customer, 'kETvYtLKZkgIzBvfRdVcZULBITANjOMd', 'sha512', 'latin
```

16.4.2. C.Decode And C.Encode - Encoding And Decoding

C.Decode – Data Decoding Methods

C.Decode.base64()

```
Decode.base64(val: string, resultEnc: string = 'utf8'): string | Buffer | undefine
```

Performs base64 decoding of the given string.

Returns a string or Buffer, depending on the `resultEnc` value.

Parameter	Type	Description
`val`	string	Value to decode.
`resultEnc`	`'utf8' | 'utf8-valid' | 'buffer'`	Encoding to use to convert the binary data to a string. Use `'utf8-valid'` to validate that result is valid UTF8; use `'buffer'` if you need the binary data in a Buffer. Defaults to `'utf8'`.

Examples

To decode an `encoded` field and return the decoded value as a Buffer, use:

```
C.Decode.base64(encoded, 'buffer')
```

C.Decode.gzip()

```
Decode.gzip(value: Buffer | string, encoding?: BufferEncoding) : string
```

Gunzips the supplied value.

Parameter	Type	Description
`value`	any	Value to gunzip.
`encoding`	string	Encoding of `value`, for example: `'base64'`, `'hex'`, `'utf-8'`, `'binary'`. Default is `'base64'`. If data is received as Buffer (from gzip with encoding: `'none'`), decoding is skipped.

Examples

To decode a `hostname` field which had been encoded using `hex`, run:

```
C.Decode.gzip(hostname, 'hex')
```

C.Decode.hex()

```
Decode.hex(val: string): number
```

Performs hex to number conversion. Returns `NaN` if value cannot be converted to a number.

Parameter	Type	Description
`val`	string	Hex string to parse to a number (for example, “0xcafe”).

C.Decode.uri()

```
Decode.uri(val: string): string
```

Performs URI-decoding of the given string.

Parameter	Type	Description
`val`	string	Value to decode.

C.Decode.inflate()

```
Decode.inflate(value: Buffer | string, encoding: BufferEncoding = 'base64', isRaw =
```

Inflates the supplied value.

Parameter	Type	Description
`value`	string | Buffer	The value to inflate. May be a Buffer or a string.
`encoding`	string	If a string `value` is provided, this specifies how it has been encoded. If no `encoding` is provided, `base64` is used. This parameter is ignored if `value` is already a Buffer.
`isRaw`	boolean	Indicates whether the data is in raw deflate format (without zlib headers). Default value is `false`.

Examples

To base64-decode and inflate the value contained in a `_raw` field, you can use:

```
C.Decode.inflate(_raw)
```

To additionally specify a different encoding to use, run:

```
C.Decode.inflate(_raw, 'hex')
```

C.Encode – Data Encoding Methods

C.Encode.base64()

```
Encode.base64(val: string | Buffer, trimTrailEq: boolean = false, encoding?: Buffe
```

Returns a base64 representation of the given string or Buffer.

Parameter	Type	Description
`val`	any	Value to encode.
`trimTrailEq`	boolean	Whether to trim any trailing `=`.
`encoding`	string	Optional parameter that specifies the encoding of the value. Options

Parameter	Type	Description
		include `base64`, `hex`, `utf-8`, and `binary`. The default is `utf-8`.

Examples

To base64-encode the `username` field, and remove the trailing `=`, use:

```
C.Encode.base64(username, true)
```

To base64-encode the `hexValue` field, where the value of the field is a hex string:

```
C.Encode.base64('4672616e6b7575757575', undefined, 'hex')
```

C.Encode.gzip()

```
Encode.gzip(value: Buffer | string, encoding?: string) : string | Buffer
```

Gzips, and optionally base64-encodes, the supplied value.

Parameter	Type	Description
`value`	string	Value to gzip.
`encoding`	string	Encoding of `value`, for example: `'base64'`, `'hex'`, `'utf-8'`, `'binary'`, `'none'`. Default is `'base64'`. If `'none'` is specified, data will be returned as a Buffer.

Examples

To encode the `hostname` field, for example using `hex` encoding, run:

```
C.Encode.gzip(hostname, 'hex')
```

C.Encode.hex()

```
Encode.hex(val: string | number): string
```

Rounds the number to an integer and returns its hex representation (lowercase). If a string is provided, it will be parsed into a number or `NaN`.

Parameter	Type	Description
`val`	string | number	Value to convert to hex.

`C.Encode.uri()`

```
Encode.uri(val: string): string
```

Returns the URI-encoded representation of the given string.

Parameter	Type	Description
`val`	string	Value to encode.

`C.Encode.deflate()`

```
Encode.deflate(value: Buffer | string, encoding: BufferEncoding | 'none' = 'base64
```

Deflates and optionally base64-encodes the supplied value.

Parameter	Type	Description
`value`	string | Buffer	Value to deflate.
`encoding`	string	Encoding to apply to the deflated result, for example: `base64`, `hex`, `utf-8`, `binary`, `none`. Default is `base64`. If `none` is specified, data will be returned as a Buffer.
`toRaw`	boolean	Indicates whether to return the raw deflated data, without zlib headers. Default is `false`, which will prepend normal zlib header bytes to the returned value.

Examples

To deflate and hex-decode the value provided in an `inflated` field, run:

```
C.Encode.deflate(inflated, 'hex')
```

16.4.3. C.Lookup – Inline Lookup Methods

C.Lookup – Exact Lookup

```
Lookup: (file: string, primaryKey?: string, otherFields: string[]=[],ignoreCase: bo
```

Returns an instance of a lookup to use inline.

Examples

In this example, `host` is the name of the primary key field:

```
C.Lookup('lookup_name.csv', 'IP_field_name_in_lookup_file').match(host)
```

Here, the quoted `'event_field_or_string_to_match'` could be a string to match in the primary key field:

```
C.Lookup('name_of_lookup_file.csv', 'field_in_csv_to_match').match('event_field_or_
```

This example checks whether `someValue` is present or not and returns the answer as a boolean.

```
C.Lookup('file', 'primaryKeyCol', additionalCols).match('someValue')
```

This expression returns a string – the `fieldToReturn` column from the matched lookup row:

```
C.Lookup('file', 'primaryKeyCol', additionalCols).match('someValue','fieldToReturn
```

This expression returns an array – the specified `fieldToReturn1` and `fieldToReturn2` columns from the matched lookup row:

```
C.Lookup('file', 'primaryKeyCol').match('someValue', ['fieldToReturn1', 'fieldToRet
```

To return an array with all columns from the matched lookup row, you can use:

```
C.Lookup('file', 'primaryKeyCol').match('someValue', [])
```

`C.Lookup` can load lookup files of up to 10 MB.

All inputs to Lookup methods' `match()` method must be strings. If your lookup file contains numeric fields, convert them to strings, e.g.: `.match(String(<fieldname>))`.

You can use the optional `otherFields[]` argument, shown in the above `C.Lookup()` signatures and examples, to limit which columns of the lookup table will be available in a subsequent `.match()` call. If omitted, or set to `undefined`, all columns will be available.

Cribl Edge 4.1 and later support returning all columns from a matched row. Use the empty-array convention shown in the above examples: `.match('someValue',[])`.

C.LookupCIDR – CIDR Lookup

```
LookupCIDR: (file: string, primaryKey?: string, otherFields: string[]=[]) => Inline
```

Returns an instance of a CIDR lookup to use inline.

C.LookupIgnoreCase – Case-insensitive Lookup

```
LookupIgnoreCase: (file: string, primaryKey?: string, otherFields: string[]=[]) =>
```

Returns an instance of a lookup (ignoring case) to use inline. Works identically to `C.Lookup`, except ignores the case of lookup values. (Equivalent to calling `C.Lookup` with its fourth `ignoreCase?` parameter set to `true`).

C.LookupRegex – Regex Lookup

```
LookupRegex: (file: string, primaryKey?: string, otherFields: string[]=[]) => Inli
```

Returns an instance of a Regex lookup to use inline.

C.Lookup.match()

```
InlineLookup.match(value: string): boolean

InlineLookup.match(value: string, fieldToReturn?: string): any
InlineLookup.match(value: string, fieldToReturn: string[]): {[key: string]: any}
```

Parameter	Type	Description
`value`	string	the value to look up.
`fieldToReturn`	string | string []	name of the lookup file > field to return.

Examples

To find the last row where the value from column `foo` matches the string `abc`, you can use the following expression. If a matching row is found, the expression returns the row's value for column `bar`.

```
C.Lookup('lookup-exact.csv', 'foo').match('abc', 'bar')
```

To find the last row where the value from column `transport` matches the string `tcp`, you can use the following expression. If a matching row is found, the expression returns the row's value for column `port_number`.

```
C.Lookup('service_names_port_numbers.csv', 'transport').match('tcp', 'port_number')
```

You can find the last row where the CIDR range in `cidr` includes `192.168.1.1` with the following expression. If a matching row is found, the expression returns the row's value for column `bar`.

```
C.LookupCIDR('lookup-cidr.csv', 'cidr').match('192.168.1.1', 'bar')
```

The following expression finds the last row where the value from column `cidr` matches `hostIP`. If a matching row is found, it returns the row's value for column `location`.

```
C.LookupCIDR('lookup-cidr.csv', 'cidr').match(hostIP, 'location')
```

To find the last row where the regex in column `foo` matches the string `manchester`, use the following expression. If a matching row is found, the expression returns the row's value for column `bar`.

```
C.LookupRegex('lookup-regex.csv', 'foo').match('manchester', 'bar')
```

With `C.LookupRegex`, ensure that your lookup file contains no empty lines – not even at the bottom. Any empty rows will cause `C.LookupRegex().match()` to always return `true`.

16.4.4. C.Mask – Data Masking Methods

C.Mask.CC()

```
Mask.CC(value: string, unmasked: number = -4, maskChar: string='X'): string
```

Checks whether a value could be a valid credit card number, and masks a subset of the value. By default, all digits except the last 4 will be replaced with `X`.

Parameter	Type	Description
`value`	string	The string to mask, **if and only if** it could be a valid credit card number.
`unmasked`	number	Number of digits to leave unmasked: positive for left, negative for right, `0` for none.
`maskChar`	string	String to replace each digit with.

Examples

Let's assume your parsed data contains a `creditCard` field. You can mask this number by using a Mask Function. Apply it to the `cardNumber` field and create the following masking rule:

Match Regex	Replace Expression	Example result
`(.*)` (catches the whole field contents)	`${C.Mask.CC(g1)}`	`cardNumber: XXXXXXXXXX4860`

To control how the masking is performed, you can specify how many digit should be unmasked and what masking character to use:

Match Regex	Replace Expression	Example result
`(.*)` (catches the whole field contents)	`${C.Mask.CC(g1, 4, "Z")}`	`cardNumber: 6011ZZZZZZZZZZZZ`

C.Mask.IMEI()

```
Mask.IMEI(value: string, unmasked: number = -4, maskChar: string='X'): string
```

Checks whether a value could be a valid IMEI number, and masks a subset of the value. By default, all digits except the last 4 will be replaced with `X`.

Parameter	Type	Description
`value`	string	The string to mask, **if and only if** it could be a valid IMEI number.
`unmasked`	number	Number of digits to leave unmasked: positive for left, negative for right, `0` for none.
`maskChar`	string	String to replace each digit with.

Examples

Let's assume your parsed data contains an `imei` field. You can mask this number by using a Mask Function. Apply it to the `imei` field and create the following masking rule:

Match Regex	Replace Expression	Example result
`(.*)` (catches the whole field contents)	`${C.Mask.IMEI(g1)}`	`imei: XXXXXXXXXXX7011`

To control how the masking is performed, you can specify how many digit should be unmasked and what masking character to use:

Match Regex	Replace Expression	Example result
`(.*)` (catches the whole field contents)	`${C.Mask.CC(g1, 4, "M")}`	`cardNumber: 5356MMMMMMMMMM`

C.Mask.isCC()

```
Mask.isCC(value: string): boolean
```

Checks whether the given value could be a valid credit card number, by computing the string's Luhn's checksum modulo 10 == `0`.

Parameter	Type	Description
`value`	string	String to check for being a valid credit card number.

Examples

The following example expression uses `isCC()` to check whether the `cardNumber` field actually contains a credit card number. If it does, the card number is replaced with the `REDACTED` string, otherwise, en empty string is returned.

```
C.Mask.isCC(cardNumber) ? C.Mask.REDACTED : ""
```

C.Mask.isIMEI()

```
Mask.isIMEI(value: string): boolean
```

Checks whether the given value could be a valid IMEI number, by computing the string's Luhn's checksum modulo 10 == `0`.

Parameter	Type	Description
`value`	string	String to check for being a valid IMEI number.

Examples

The following example expression uses `isCC()` to check whether the `imei` field actually contains ab IMEI number number. If it does, the IMEI number is replaced with the `REDACTED` string, otherwise, en empty string is returned.

```
C.Mask.isIMEI(imei) ? C.Mask.REDACTED : ""
```

C.Mask.luhn()

```
Mask.luhn(value: string, unmasked: number = -4, maskChar: string='X'): string
```

Checks that value Luhn's checksum mod 10 is `0`, and masks a subset of the value. By default, all digits except the last 4 will be replaced with `X`. If the value's Luhn's checksum mod 10 is not `0`, then the value is returned unmodified.

Parameter	Type	Description
`value`	string	The string to mask, **if and only if** the value's Luhn's checksum mod 10 is `0`.
`unmasked`	number	Number of digits to leave unmasked: positive for left, negative for right, `0` for none.
`maskChar`	string	String to replace each digit with.

C.Mask.luhnChecksum()

```
Mask.luhnChecksum(value: string, mod: number=10): number
```

Generates the Luhn checksum (used to validate certain credit card numbers, IMEIs, etc.). By default, the mod 10 of the checksum is returned. Pass mod = `0` to get the actual checksum.

Parameter	Type	Description
`value`	string	String whose digits you want to perform the Luhn checksum on.
`mod`	number	Return checksum modulo this number. If `0`, skip modulo. Default is `10`.

C.Mask.md5()

```
Mask.md5(value: string, len?: string | number, encoding?: BufferEncoding): string
```

Generates MD5 hash of a given value.

Parameter	Type	Description
`value`	string	The string to hash.
`len`	string | number	Length of hash to return: 0 for full hash, a +number for left or a -number for right substring. If a string is passed it's length will be used.

Parameter	Type	Description
`encoding`	string	Optional parameter that specifies the encoding of the value. Options include `base64`, `hex`, `utf-8`, and `binary`. The default is `utf-8`.

Examples

To hash the contents of a `username` field with the MD5 algorithm, run:

```
C.Mask.md5(username)
```

To hash the contents of an `authToken` field, where the value of the field is a base64-encoded string, with the MD5 algorithm, run:

```
C.Mask.md5(authToken, undefined, 'base64')
```

C.Mask.random

```
Mask.random(len?: string | number): string
```

Generates a random alphanumeric string.

Parameter	Type	Description
`len`	string | number	Length of the result; or, if a string, use its length.

Examples

Assuming you have some identifying user information in a `username` field, you can replace it with a random string of the same length with:

```
C.Mask.random(username)
```

Input	Output
`McGoatFace`	“xY2KLHkyTg”

C.Mask.repeat()

```
Mask.repeat(len?: number | string, char: string = 'X'): string
```

Generates a repeating char/string pattern, e.g., XXXX.

Parameter	Type	Description
len	string | number	Length of the result; or, if a string, use its length.
char	string	Pattern to repeat len times.

Examples

The following example will repeat the “GOAT” string as many time as there are characters in the first parameter, “goat”, so four times:

```
C.Mask.repeat("goat", "GOAT")
```

C.Mask.sha1(), C.Mask.sha256(), C.Mask.sha512()

```
Mask.sha1(value: string, len?: string | number, encoding?: BufferEncoding): string
Mask.sha256(value: string, len?: string | number, encoding?: BufferEncoding): strir
Mask.sha512(value: string, len?: string | number, encoding?: BufferEncoding): strir
```

Generates SHA1, SHA256, or SHA512 hash of given value.

Parameter	Type	Description
value	string	The string to hash.
len	string | number	Length of hash to return: 0 for full hash, a +number for left, or a -number for right substring. If a string is passed, its length will be used.
encoding	string	Optional parameter that specifies the encoding of the value. Options include base64, hex, utf-8, and binary. The default is utf-8.

Examples

To hash the contents of a `secret` field with the SHA256 algorithm, run:

```
C.Mask.sha256(secret)
```

To hash the contents of an `authToken` field, where the value of the field is a base64-encoded string, with the SHA256 algorithm, run:

```
C.Mask.sha256(authToken, undefined, 'base64')
```

C.Mask.REDACTED

```
Mask.REDACTED: string
```

The literal `'REDACTED'`.

16.4.5. C.Net – Network Methods

C.Net.cidrMatch()

```
Net.cidrMatch(cidrIpRange: string, ipAddress: string): boolean
```

Determines whether the supplied IPv4 or IPv6 `ipAddress` is inside the range of addresses identified by `cidrIpRange`.

Parameter	Type	Description
`cidrIpRange`	string	IP address range in CIDR format. E.g., `10.0.0.0/24`.
`ipAddress`	string	The IP address to test for inclusion in `cidrIpRange`.

Examples

(IPv4): `C.Net.cidrMatch('10.0.0.0/24', '10.0.0.100')` returns `true`.

(IPv6): `C.Net.cidrMatch('2001:db8::/32', '2001:db8:0:0:0:ffff:0:0')` returns `true`.

C.Net.isIpV4()

```
Net.isIpV4(value: string): boolean
```

Determines if the value supplied is an IPv4 address.

Parameter	Type	Description
`value`	string	The IP address to test.

C.Net.isIpV6()

```
Net.isIpV6(value: string): boolean
```

Determines if the value supplied is an IPv6 address.

Parameter	Type	Description
`value`	string	The IP address to test.

C.Net.isIPV4Cidr()

```
Net.isIPV4Cidr(value: string): boolean
```

Determines whether a value supplied is an IPv4 CIDR range.

Returns `true` if the value is a valid IPv4 CIDR range; otherwise, `false`.

Parameter	Type	Description
`value`	string	String value that may be an IPv4 CIDR range.

C.Net.isIPV6Cidr()

```
Net.isIPV6Cidr(value: string): boolean
```

Determine whether a value supplied is an IPv6 CIDR range.

Returns `true` if the value is a valid IPv6 CIDR range; otherwise, `false`.

Parameter	Type	Description
`value`	string	String value that may be an IPv6 CIDR range.

C.Net.isIpV4AllInterfaces()

```
Net.isIpV4AllInterfaces(value): boolean
```

Determines if the value supplied is the IPv4 all-interfaces address, typically used to bind to all IPv4 interfaces – i.e., '0.0.0.0'.

Parameter	Type	Description
`value`	any	The IP address to test.

C.Net.isIpV6AllInterfaces()

```
Net.isIpV6AllInterfaces(value): boolean
```

Determines if the value supplied is an IPv6 all-interfaces address, typically used to bind to all IPv6 interfaces – one of: '::', '0:0:0:0:0:0:0:0', or '0000:0000:0000:0000:0000:0000:0000:0000'.

Parameter	Type	Description
`value`	any	the IP address to test.

C.Net.ipv6Normalize()

```
Net.ipv6Normalize(address: string): string
```

Normalizes an IPV6 address based on RFC draft-ietf-6man-text-addr-representation-04.

Parameter	Type	Description
`address`	string	The IPV6 address to normalize.

C.Net.isPrivate()

```
Net.isPrivate(address: string): boolean
```

Determines whether the supplied IPv4 address is in the range of private addresses per RFC1918.

Parameter	Type	Description
`address`	string	The IP address to test.

16.4.6. C.Text – Text Methods

C.Text.entropy()

```
Text.entropy(bytes: Buffer | string): number
```

Computes the Shannon entropy of the given Buffer or string.

Returns the entropy value; or `-1` in case of an error.

Parameter	Type	Description
`bytes`	Buffer | string	value to undergo Shannon entropy computation.

C.Text.hashCode()

```
Text.hashCode(val: string | Buffer | number): number
```

Computes hashcode (djb2) of the given value.

Returns hashcode value.

Parameter	Type	Description
`val`	string | Buffer | number	value to be hashed.

C.Text.isASCII()

```
Text.isASCII(bytes: Buffer | string): boolean
```

Checks whether all bytes or chars are in the ASCII printable range.

Returns `true` if all chars/bytes are within ASCII printable range; otherwise, `false`.

Parameter	Type	Description
`bytes`	string | Buffer	value to check for character range.

C.Text.isUTF8()

```
Text.isUTF8(bytes: Buffer | string): boolean
```

Checks whether the given Buffer contains valid UTF8.

Returns `true` if bytes are UTF8; otherwise, `false`.

Parameter	Type	Description
`bytes`	Buffer | string	bytes to check.

C.Text.parseWinEvent()

```
Text.parseWinEvent(xml: string, nonValues: string[] = Text._WIN_EVENT_NON_VALUES):
```

Parses an XML string representing a Windows event into a compact, prettified JSON object. Works like `C.Text.parseXml`, but with Windows events, produces more-compact output. For a usage example, see Reducing Windows XML Events.

Returns an object representing the parsed Windows Event; or `undefined` if the input could not be parsed.

Parameter	Type	Description
`xml`	string	an XML string; or an event field containing the XML.
`nonValues`	string[]	array of string values. Elements whose value equals any of the values in this array will be omitted from the returned object. Defaults to `['-']`, meaning that elements whose value equals `-` will be discarded.

C.Text.parseXml()

```
Text.parseXml(xml: string, keepAttr: boolean = true, keepMetadata: boolean = false
```

Parses an XML string and returns a JSON object. Can be used with Eval Function to parse XML fields contained in an event, or with ad hoc XML.

Returns an object representing the parsed XML; or `undefined` if the input could not be parsed. An input collection of elements will be parsed into an array of objects.

Parameter	Type	Description
`xml`	string	XML string, or an event field containing the XML.
`keepAttr`	boolean	whether or not to include attributes in the returned object. Defaults to `true`.
`keepMetadata`	boolean	whether or not to include metadata found in the XML. The `keepAttr` parameter must be set to `true` for this to work. Defaults to `false`. (Eligible metadata includes namespace definitions and prefixes, and XML declaration attributes such as encoding, version, etc.)
`nonValues`	string[]	array of string values. Elements whose value equals any of the values in this array will be omitted from the returned object. Defaults to `[]` (empty array), meaning discard no elements.

C.Text.relativeEntropy()

```
Text.relativeEntropy(bytes: Buffer | string, modelName: string = 'top_domains'): nu
```

Computes the relative entropy of the given Buffer or string.

Returns the relative entropy value, or `-1` in case of an error.

Parameter	Type	Description
`bytes`	Buffer | string	Value whose relative entropy to compute.
`modelName`	string	Optionally, override the default `$CRIBL_HOME/data/lookups/model_relative_entropy_top_domains.csv` model used to test the input. Create a custom lookup file with the same column and value structure as the default, and store it in the same path, as `model_relative_entropy_<custom-name>.csv`. To reference it, pass your `<custom-name>` substring as the `modelName` parameter.

When using `modelName` in a distributed deployment, the corresponding paths are `$CRIBL_HOME/groups/<worker-group-name>/data/lookups/`. Creating your custom lookup file via Cribl Edge's UI will automatically set the appropriate paths.

16.4.7. C.Time – Time Methods

C.Time.adjustTZ()

```
Time.adjustTZ(epochTime: number, tzTo: string, tzFrom: string = 'utc'): number | ur
```

Adjusts a timestamp from one timezone to another.

Returns the adjusted timestamp, in UNIX epoch time (ms).

Parameter	Type	Description
`epochTime`	number	Timestamp to adjust, in UNIX epoch time.
`tzTo`	string	Timezone to adjust to (full name).
`tzFrom` (optional)	string	Timezone of the timestamp.

Cribl relies on this list of TZ Database Time Zones.

Note that these are timezone identifiers (full names), not abbreviations such as "CET", "PST", "EST", and so on.

Examples

To adjust a timestamp found in the `_time` field to a selected timezone (here, Australia/Melbourne), use:

```
C.Time.adjustTZ(_time, "Australia/Melbourne")
```

The method returns the timestamp in milliseconds. If you want to further operate on it using methods such as `strftime()`, which takes as parameter UNIX epoch time in seconds, you can divide it by 1000:

```
C.Time.strftime(C.Time.adjustTZ(_time, "Australia/Melbourne") / 1000, "%d %B %Y %H
```

C.Time.clamp()

```
Time.clamp(date: T, earliest: T, latest: T, defaultDate?: T ): T  | undefined
```

Constrains a parsed timestamp to realistic earliest and latest boundaries.

Returns a JavaScript Date object or timestamp, depending on the format of the `date` parameter.

Parameter	Type	Description
`date`	Date | number	Timestamp to constrain, in UNIX epoch time (ms) or JavaScript Date format.
`earliest`	Date | number	Earliest allowable timestamp, in UNIX epoch time (ms) or JavaScript Date format.
`latest`	Date | number	Latest allowable timestamp, in UNIX epoch time (ms) or JavaScript Date format.
`defaultDate` (optional)	Date | number	Default date, in UNIX epoch time (ms) or JavaScript Date format. This date is used for values that fall outside the `earliest` or `latest` boundaries.

Examples

Suppose you have a date in a "YYYY-MM-DD" format in a `time` field. To make sure this date does not exceed the range of 1 Jan 2020 – 1 Jan 2025, you can use:

```
C.Time.clamp(time, "2020-01-01", "2025-01-01")
```

Input	Output
"2016-02-11"	"2020-01-01" (set to the earliest value)
"2026-01-01"	"2025-01-01" (set to the latest value)
"2023-12-24"	"2023-12-24" (unchanged, because it fits inside the defined range)

To make sure that all dates that fall outside the bounds are changed to a specific value, add the third parameter:

```
C.Time.clamp(time, "2020-01-01", "2025-01-01", "2022-02-22")
```

Input	Output
“2016-02-11”	“2022-02-22” (set to the default value)
“2026-01-01”	“2022-02-22” (set to the default value)
“2023-12-24”	“2023-12-24” (unchanged, because it fits inside the defined range)

C.Time.strftime()

```
Time.strftime(date: number | Date | string, format: string, utc: boolean = true): s
```

Formats a Date object or timestamp number as a time string, using d3js time format.

Returns representation of the given date in the specified format.

Parameter	Type	Description
`date`	number | Date | string	Date to format, in UNIX epoch time (in seconds) or JavaScript Date format.
`format`	string	New format for the date.
`utc` (optional)	boolean	Whether to output the time in UTC (`true`), rather than in local timezone.

Examples

To transform a `_time` field containing a UNIX timestamp to a string with time formatted, for example: `15 November 2023, 08:03`, you can use:

```
C.Time.strftime(_time, "%d %B %Y, %H:%M")
```

Some common formats (assuming UNIX timestamp of 1702460134) are:

Format	Output
“%Y-%m-%d” (ISO date)	“2023-12-13”
“%d %B %Y %H:%M:%S”	“13 December 2023 09:35:34”

Format	Output
“%x %X” (the locale’s date and time format)	“12/13/2023 9:35:34 AM”

For detailed reference of the date and time tokens, see d3js reference or Auto Timestamp.

C.Time.strptime()

```
Time.strptime(str: string, format: string, utc: boolean = true, strict: boolean =
```

Extracts time from a string using d3js time format that defines the format of the incoming string.

Returns a parsed JavaScript Date object, or `null` if the specified `format` did not match.

Parameter	Type	Description
`str`	string	Timestamp to parse.
`format`	string	Format to parse.
`utc` (optional)	boolean	Whether to interpret times as UTC (`true`), rather than as local time.
`strict` (optional)	boolean	Whether to return `null` if there are any extra characters after timestamp.

Examples

To convert a string in the “YYYY-MM-DD” format, for example “2023-12-12”, to a Date object, use:

```
C.Time.strptime(F, "%Y %m %d")
```

Additionally, to interpret the time in UTC, and make sure null is returned if the timestamp contains any trailing characters, use:

```
C.Time.strptime(F, "%Y %m %d", true, true)
```

For detailed reference of the date and time tokens, see d3js reference or Auto Timestamp.

C.Time.timestampFinder()

```
Time.timestampFinder(utc?: boolean).find(str: string): IAutoTimeParser
```

Extracts time from the specified field, using the same algorithm as the Auto Timestamp Function and the Event Breaker Function.

Returns representation of the extracted time; truncates timestamps to three-digit (milliseconds) resolution, omitting trailing zeros.

Parameter	Type	Description
`utc`	boolean	Whether to output the time in UTC (`true`), rather than in local timezone.
`str`	string	The field in which to search for the time.

Examples

To find timestamps in the `_raw` field and output them in UTC, you can use:

```
C.Time.timestampFinder(true).find(_raw)
```

You can then format the string output with `strftime()`:

```
C.Time.strftime(C.Time.timestampFinder(true).find(_raw), "%d.%m.%Y, %H.%M")
```

16.4.8. Miscellaneous Expression Methods

C.Schema – Schema Methods

C.Schema

```
Schema: (id: string) => SchemaValidator
```

C.Schema.validate()

```
SchemaValidator.validate(data: any): boolean
```

Validates the given object against the schema.

Returns `true` when schema is valid; otherwise, `false`.

Parameter	Type	Description
`data`	any	Object to be validated.

Examples

To validate whether a `myField` conforms to `schema1`, you can use:

```
C.Schema('schema1').validate(myField)
```

See Schema Library for more details.

C.Secret – Secrets-Management Methods

C.Secret()

```
Secret: (id: string, type?: string): ISecret
Secret(id: string, type: 'keypair') => IPairSecret
Secret(id: string, type: 'text') => ITextSecret
Secret(id: string, type: 'credentials') => ICredentialsSecret
```

Returns a secret matching the specified ID.

Parameter	Type	Description
`id`	string	ID of the secret.
`type` (optional)	'text' | 'keypair' | 'credentials'	Type of the secret.

Examples

To return a text secret with ID `victorias` (or with undefined, if no such secret exists), use:

```
C.Secret('victorias', 'text')
```

You can also get attributes of secrets with the following expressions:

```
C.Secret('api_key', 'keypair').secretKey
C.Secret('secret_hash', 'text').value
C.Secret('user_pass', 'credentials').password
```

Common returned attributes for `ISecret` objects:

- `secretType` – one of `keypair`, `text`, or `credentials`.
- `description` (optional) – the secret description.
- `tags` (optional) – a comma separated list of tags.

Additional returned attributes for `IPairSecret` objects:

- `apiKey` – the API key value
- `secretKey` – the Secret key value

Additional returned attributes for `ITextSecret` objects:

- `value` – the text value

Additional returned attributes for `ICredentialsSecret` objects:

- `username` – the username value
- `password` – the password value

See Securing Cribl Stream > Secrets for more details.

C.env – Environment

`C.env`

```
env: {[key: string]: string;}
```

Returns an object containing Cribl Edge's environment variables.

Examples

To return the parent of Cribl's `bin` directory (generally `/opt/cribl`), use:

```
C.env.CRIBL_HOME
```

To return the hostname of the machine where Cribl Edge is running, use:

```
C.env.HOSTNAME
```

C.os – System Methods

`C.os.hostname()`

Returns hostname of the system running this Cribl Edge instance.

C.vars – Global Variables

See Global Variables Library for more details.

C.version – Cribl Edge Versions

C.version

Returns the Cribl Edge version currently running.

C.confVersion

Returns the commit hash of the Edge Node's current config version. (Evaluates only against live data sent through Edge Nodes. Values will be undefined in the Leader's Preview pane.)

C.Misc – Miscellaneous Utility Methods

C.Misc.zip()

```
Misc.zip(keys: string[], values: any[], dest?: any): any
```

Sets the given keys to the corresponding values on the given `dest` object. If `dest` is not provided, a new object will be constructed.

Returns object on which the fields were set.

Parameter	Type	Description
`keys`	string[]	Field names corresponding to keys.
`values`	any[]	Values corresponding to values.
`dest`	any	Object on which to set field values.

Examples

Let's take the following expression:

```
people = C.Misc.zip(titles, names)
```

If sample data contains: `titles=['ceo', 'svp', 'vp']`, `names=['foo', 'bar', 'baz']`, this expression create an object called `people`, with key names from elements in `titles`, and with

corresponding values from elements in `names`.

Result: `"people": {"ceo": "foo", "svp": "bar", "vp": "baz"}`

C.Misc.uuidv4()

```
C.Misc.uuidv4(): string
```

Returns a version 4 (random) UUID in accordance with RFC-4122.

Examples

To create a UUIDv4, use:

```
C.Misc.uuidv4()
```

Result: a conforming UUIDv4, such as `58d8be36-4db0-4b1c-ac80-28bb03c45e0d`. It is highly improbable that two version 4 UUIDs will ever have the same value.

C.Misc.uuidv5()

```
C.Misc.uuidv5(name: string, namespace: string): string
```

Returns a version 5 (namespaced) UUID in accordance with RFC-4122 for the given `name` and `namespace`. If `namespace` is not a valid UUID, this function will fail.

Parameter	Type	Description
`name`	string	Any arbitrary name to use in UUID generation.
`namespace`	string	One of `DNS`, `URL`, `OID`, or `X500` to use a predefined namespace, or else a valid UUID.

Examples

To create a UUIDv5 with the predefined DNS namespace, use:

```
C.Misc.uuidv5('example', 'DNS')
```

Result: a UUID that is highly likely to be the same for the same name and namespace. In this case, the result would be `7cb48787-6d91-5b9f-bc60-f30298ea5736`.

C.Misc.validateUUID()

```
C.Misc.validateUUID(maybeUUID: string): boolean
```

Returns `true` if `maybeUUID` is a valid UUID of any version, and false otherwise.

Parameter	Type	Description
`maybeUUID`	string	A string to test for a valid UUID.

Examples

```
C.Misc.validateUUID(C.Misc.uuidv4())
```

Result: returns `true`

```
C.Misc.validateUUID('clearly not a UUID')
```

Result: returns `false`

C.Misc.getUUIDVersion()

```
C.Misc.getUUIDVersion(uuid: string): number
```

Returns a `number` of the UUID version given by `uuid` if it is a valid UUID, otherwise `undefined`.

Parameter	Type	Description
`uuid`	string	A UUID for which to determine the version.

Examples

```
C.Misc.getUUIDVersion(C.Misc.uuidv4())
```

Result: `4`

```
C.Misc.getUUIDVersion(C.Misc.uuidv5('example', 'X500'))
```

Result: `5`

16.5. Cribl Expression Syntax

As data travels through a Cribl Edge Pipeline, it is operated on by a series of Functions. Functions are fundamentally JavaScript code.

Functions that ship with Cribl Edge are configurable via a set of inputs. Some of these configuration options are literals, such as field names, and others can be JavaScript expressions.

Expressions are **valid units** of code that resolve to a value. Every syntactically valid expression resolves to some value, but conceptually, there are two types of expressions: those that **assign** value to a variable (a.k.a., with side effects), and those that **evaluate** to a value.

Assigning a value	Evaluating to a value

`x = 42`
`newFoo = foo.slice(30)`
	`(Math.random() * 42)`
	`3 + 4`
	`'foobar'`
	`'42'`

Filters and Value Expressions

Let's consider evaluation expressions before assignment expressions.

Filters

Filters are used in Routes to select a stream of the data flow, and in Functions to scope or narrow down the applicability of a Function. Filters are expressions that **must** evaluate to either `true` (or truthy) or `false` (or falsy). Keep this in mind when creating Routes or Functions. For example:

- `sourcetype=='access_combined' && host.startsWith('web')`
- `source.endsWith('.log') || sourcetype=='aws:cloudwatchlogs:vpcflow'`

This table shows examples of truthy and falsy values.

Truthy	Falsy

`true`
`42`
`-42`
`3.14`
`"foo"`
	`false`
	`null`
	`undefined`
	`0`
	`NaN`

Truthy	Falsy
`Infinity` `-Infinity`	`''` `""`

Value Expressions

Value expressions are typically used in Functions to assign a value – for example, to a new field. For example:

- `Math.floor(_time/3600)`
- `source.replace(/.{3}/, 'XXX')`

Best Practices for Creating Predictable Expressions

- In a value expression, ensure that the source variable is not `null`, `undefined`, or `empty`.
  For example, assume you want to have a field called `len`, to be assigned the length of a second field called `employeeID`. But you're not sure if `employeeID` exists. Instead of `employeeID.length`, you can use a safer shorthand, such as: `(employeeID || '').length`.
- If a field does not exist (undefined), and you're doing a comparison with its properties, then the boolean expression will **always** evaluate to false. For example, if `employeeID` is undefined, then both of these expressions will evaluate to false: `employeeID.length > 10` , and `employeeID.length < 10` .
- `==` means "equal to," while `===` means "equal value **and** equal type." For example, `5 == 5` and `5 == "5"` each evaluate to **true**, while `5 === "5"` evaluates to **false**.
- A ternary operator is a very powerful way to create conditional values. For example, if you wanted to assign either `minor` or `adult` to a field `groupAge`, based on the value of `age`, you could do: `(age >= 18) ? 'adult' : 'minor'`.

Fields with Non-Alphanumeric Characters

If there are fields whose names include non-alphanumeric characters – e.g., `@timestamp` or `user-agent` or `kubernetes.namespace_name` – you can access them using `__e['<field-name-here>']`. (Note the single quotes.) More details here.

In any other place where the field is referenced – e.g., in the Eval function's field names – you should use a single-quoted literal, of the form: `'<field-name-here>'`.

Wildcard Lists

Wildcard Lists are used throughout the product, especially in various Functions, such as Eval, Mask, Publish Metrics, Parser, etc.

Wildcard Lists, as their name implies, accept strings with asterisks (`*`) to represent one or more terms. They also accept strings that start with an exclamation mark (`!`) to **negate** one or more terms. This allows for implementing any combination of allowlists and blocklists.

Wildcard Lists are order-sensitive, evaluated from left to right. This is especially relevant when you use negated terms. For negations to take precedence over wildcards when evaluated, you must list negations before wildcards.

Some examples:

Wildcard List	Value	Meaning
List 1	`!foobar, foo*`	All terms that start with **foo**, except **foobar**.
List 2	`!foo*, *`	All terms, except for those that start with **foo**.
List 3	`*, !foo`	All terms (wildcard matches first, negation isn't evaluated).

You cannot use wildcards to target Cribl Edge internal fields that start with `__` (double underscore). You must specify these fields individually. For example, `__foobartab` cannot be removed by specifying `__foo*`.

16.6. Environment Variables

This is a consolidated list of environment variables available to configure Cribl Edge instances.

Distributed Deployment

You can use the following environment variables to configure your distributed Cribl Edge instance.

Name	Purpose
`CRIBL_DIST_MASTER_URL`	URL of the Leader Node. Example: `CRIBL_DIST_MASTER_URL=tls://<authToken>@leader:42(` See Formatting Notes below.
`CRIBL_DIST_MODE`	`worker` or `master`. Defaults to `worker`, if (and only if) `CRIBL_DIST_MASTER_URL` is present.
`CRIBL_HOME`	Auto setup on startup. Defaults to parent of `bin` directory.
`CRIBL_CONF_DIR`	Auto setup on startup. Defaults to parent of `bin` directory.
`CRIBL_NOAUTH`	Disables authentication. Careful here!!
`CRIBL_TMP_DIR`	Defines the root of a temporary directory. See Formatting Notes below.
`CRIBL_VOLUME_DIR`	Sets a directory that persists modified data between different containers (ephemeral instances. When set, this environment variable overrides $CRIBL_HOME. It also creates predefined folders in the specified directory that directory already contains folders with those names, they will be overwritten.
`CRIBL_DIST_WORKER_PROXY`	Communicate to the Leader Node via a SOCKS proxy.See Formatting Note below.
`CRIBL_BOOTSTRAP`	Quickstart a Cribl instance by configuring this variable.
`CRIBL_BOOTSTRAP_HOST`	Host name for connecting to the Leader Node when setting up a new Edge Node. This variable is not related to `CRIBL_BOOTSTRAP`.
`CRIBL_USERNAME`	Used to log in or out of Cribl.
`CRIBL_PASSWORD`	Used to log in or out of Cribl.
`CRIBL_HOST`	The Host URL for authentication. Example: `CRIBL_HOST=<url> CRIBL_USERNAME=<username> CRIBL_PASSWORD=<password>`

Name	Purpose
	`$CRIBL_HOME/bin/cribl auth login`

Usage Notes

This section explains how to use certain complex environment variables.

CRIBL_DIST_MASTER_URL

Use this format:

```
<tls|tcp>://<authToken>@host:port?group=defaultGroup&tag=tag1&tag=tag2&tls.
<tls_settings>
```

Here are the components:

- `group` – The preferred Fleet assignment.
- `resiliency` – The preferred Leader failover mode.
- `volume` – The location of the NFS directory to support Leader failover.
- `tag` – A list of tags that you can use to assign (Stream, Edge) the Worker to a Fleet.
- `tls.privKeyPath` – Private Key Path.
- `tls.passphrase` – Key Passphrase.
- `tls.caPath` – CA Certificate Path.
- `tls.certPath` – Certificate Path.
- `tls.rejectUnauthorized` – Validate Client Certs. Boolean, defaults to `false`.
- `tls.requestCert` – Authenticate Client (mutual auth). Boolean, defaults to `false`.
- `tls.commonNameRegex` – Regex matching peer certificate > subject > common names allowed to connect. Used only if `tls.requestCert` is set to `true`.

To generate a random authentication token, leave `<authToken>` unchanged. You can define it to add your own token instead, but make sure it's secure enough.

CRIBL_TMP_DIR

Sources use this variable to construct temporary directories in which to stage downloaded Parquet data. If `CRIBL_TMP_DIR` is not set (the default), Cribl applications create subdirectories within your operating

system's default temporary directory:

- For Cribl Stream: `<OS_default_temporary_directory>/stream/`.
- For Cribl Edge: `<OS_default_temporary_directory>/edge/`.

For example, on Linux, Stream's default staging directory would be `/tmp/stream/`.

If you explicitly set this `CRIBL_TMP_DIR` environment variable, its value replaces this OS-specific default parent directory.

CRIBL_DIST_WORKER_PROXY

Use the format `<socks4|socks5>://<username>:<password>@<host>:<port>`. Only `<host>:<port>` are required.

The default protocol is `socks5://`, but you can specify `socks4://proxyhost:port` if needed.

To authenticate on a SOCKS4 proxy with username and password, use this format: `username:password@proxyhost:port`. The `proxyhost` can be a `hostname`, `ip4`, or `ip6`.

CRIBL_BOOTSTRAP_HOST

`CRIBL_BOOTSTRAP_HOST` overrides the Add/Bootstrap New Worker script generator's default hostname.

For example, if you set `CRIBL_BOOTSTRAP=myhost`, then `myhost` will appear in the script modal's **Leader hostname/IP** field, instead of the URL used in the browser.

CRIBL_BOOTSTRAP

`CRIBL_BOOTSTRAP` enables specifying a URL, an absolute disk file path, or a YAML string, in order to bootstrap a configuration to the `$CRIBL_HOME/local` directory. Cribl Edge applies this configuration only upon its first startup.

For any method, Cribl Edge expects each targeted config file to be YAML-formatted. Each file's top-level keys should be the paths to config files inside the `$CRIBL_HOME/local/...` subdirectory.

Below is an example of a bootstrap file. Its output, when Cribl Edge starts, would be to create three files inside the `$CRIBL_HOME/local/cribl` path: `inputs.yml`, `outputs.yml`, and `pipelines/route.yml`.

```
cribl/inputs.yml:
  inputs:
    <id>:
      <config>
cribl/outputs.yml:
  outputs:
    <id>:
      <config>
cribl/pipelines/route.yml:
  id: default
  groups: {}
  comments: []
  routes:
    ...
```

For details about each file's syntax, see Config Files and its child topics.

Adding a Second Leader Node

You can configure a second Leader Node via the following environment variables.

Name	Purpose
`CRIBL_DIST_MASTER_RESILIENCY=failover`	Sets the Leader's `Resiliency` to `Failover`
`CRIBL_DIST_MASTER_FAILOVER_VOLUME=/tmp/shared`	Sets the location of the NFS directory to supp failover.
`CRIBL_DIST_MASTER_FAILOVER_MISSED_HB_LIMIT`	Determines how many Lease refresh periods the standby Nodes attempt to promote them primary. Cribl recommends setting this to `3`.
`CRIBL_DIST_MASTER_FAILOVER_PERIOD`	Determines how often the primary Leader ref on the Lease file. Cribl recommends setting th
`CRIBL_INSTANCE_HOME`	In `Failover` mode, this variable points to th `root` directory, as opposed to the shared vol to access `$CRIBL_INSTANCE_HOME/local/_system/` (`C:\Program Data\Cribl\local\_syster` Edge on Windows). Outside of `Failover` mo same value as `CRIBL_CONF_DIR`.

GitOps

Cribl Edge provides the following environment variables to facilitate GitOps.

Bootstrap Variables

Name	Purpose
`CRIBL_GIT_REMOTE`	Location of the remote repo to track. Can contain username and password for HTTPS auth.
`GIT_SSH / GIT_SSH_COMMAND`	See Git's documentation.
`CRIBL_GIT_BRANCH`	Git ref (branch, tag, commit) to track/check out.
`CRIBL_GIT_AUTH`	One of: `none`, `basic`, or `ssh`.
`CRIBL_GIT_USER`	Used for `basic` auth.
`CRIBL_GIT_PASSWORD`	Used for `basic` auth.
`CRIBL_GIT_OPS`	One of: `push` to enable the GitOps push workflow, or `none` to disable GitOps.
`CRIBL_GIT_SSH_KEY`	Content of the SSH key used to access git remote.
`CRIBL_GIT_STRICT_HOST_KEY_CHECKING`	Boolean flag sets whether to check the host key strictly.
`CRIBL_INTERACTIVE`	Controls whether git commands called by Cribl CLI at startup are interactive.

Internal Environment Variables

Cribl Edge uses the following variables internally.

Name	Purpose
`CRIBL_AUTO_PORTS`	When set to `true`, allows the Cribl process to listen to the first open port, if the designated API port is taken.
`CRIBL_EDGE`	When set to any value, runs this command at container start: `cribl mode-edge -H 0.0.0.0`. This launches the instance as an Edge Node, listening on a Host at `0.0.0.0`.

Name	Purpose
`CRIBL_EDGE_FS_ROOT`	Location of the host OS filesystem when mounting in a container. Defaults to `/hostfs`.
`CRIBL_WORKER_ID`	Passed to Worker processes.
`CRIBL_GROUP_ID`	Passed to ConfigHelper processes to identify Fleets.
`CRIBL_K8S_FOOTGUN`	Set to `true` to enable resource-intensive, potentially risky modes of the Kubernetes Metrics and Kubernetes Logs Sources.
`CRIBL_K8S_POD`	Sets the name of the Kubernetes Pod in which Cribl Edge is deployed.
`CRIBL_K8S_TLS_REJECT_UNAUTHORIZED`	Set to `0` to disable certification validation when connecting to the Kubernetes APIs. When you disable this environment variable, all Kubernetes features (including Metadata, Metrics, Logs, and AppScope metadata) will tolerate invalid TLS certificates (i.e., expired, self-signed, etc.) when connecting to the Kubernetes APIs.
`CRIBL_ROLE`	Controls the behavior of a Cribl subprocess, e.g., `LEADER`, `WORKER`, `CONFIG_HELPER`.
`CRIBL_SERVICE`	Set to 1 when using **systemd** to start Cribl at boot time.
`CRIBL_SERVICE_NAME`	Set to `cribl` when using **systemd** to start Cribl at boot time.
`CRIBL_SERVICEACCOUNT_PATH`	Path to the `ServiceAccount` to use to query the Kubernetes API. Defaults to `/var/run/secrets/kubernetes.io/serviceaccount`.
`CRIBL_SPOOL_DIR`	Specifies the base path where events from various Sources and Destinations are spooled. Defaults to `$CRIBL_HOME/state/spool` or `$CRIBL_VOLUME_DIR/state/spool`.

16.7. Linux System Metrics Details

Events generated by the System Metrics Source have metrics metadata to designate dimension and metric fields. The `host` field contains the hostname, and is included as a dimension in all of them. The collectors include:

- System
- CPU
- Memory
- Network
- Disk
- Container
- Process

In the Source's configuration modal, you can set the level of detail for each type of metrics:

- **Basic** enables minimal metrics, averaged or aggregated.
- **All** enables full, detailed metrics, specified for individual CPUs, interfaces, and so on.
- **Custom** displays sub-menus and buttons from which you can choose a level of detail (Basic, All, Custom, or Disabled) for each type of event.
- **Disabled** means that no metrics will be generated.

Basic and **Custom** have different meanings depending on event type and will be described under each section below.

The tables outline the metrics emitted for each mode (Basic or Custom) and where applicable, the dimensions (to indicate where the metrics are coming from).

System

With System Metrics enabled, Cribl Edge captures CPU load averages, uptime, and count. The **Custom** option allows you to include process metrics that reflect the numbers of processes in various states.

Metrics for the overall system include the following:

Name	Description	Type	Dimensions	Mode
`node_uname_info`	Labeled system information as	Counter	`release`, `sysname`,	Basic

Name	Description	Type	Dimensions	Mode
	provided by the uname system call.		`version`	
`node_cpu_count`	The number of CPU cores.	Gauge	`release`, `sysname`, `version`	Basic
`node_uptime_seconds`	System uptime in seconds.	Counter	N/A	Basic
`node_boot_time_seconds`	Node boot time in Unix time.	Counter	N/A	Basic
`node_time_seconds`	System time in seconds.	Counter	N/A	Basic
`node_load1`	1m load average.	Gauge	N/A	Basic
`node_load5`	5m load average.	Gauge	N/A	Basic
`node_load15`	15m load average.	Gauge	N/A	Basic
`node_open_fds`	Open file descriptors	Counter	N/A	Basic
`node_processes_state_all`	Total number of processes in different states.	Gauge	`state`	Basic
`node_processes_threads`	Allocated threads in system.	Gauge	`state`	Basic
`node_procs_blocked`	Number of processes blocked waiting for I/O to complete.	Gauge	`state`	Basic
`node_procs_running`	Number of processes in runnable state.	Gauge	`state`	Basic
`node_processes_state`	Number of processes in each state.	Gauge	`state`	Custom: `Process metrics`

CPU

Cribl Edge captures active, user, system, idle, and `iowait` percentages over all CPUs, with options to add per-CPU metrics and raw time counters for each state.

Metrics for CPUs include the following:

Name	Description	Type	Dimensions	Mode
`node_cpu_percent_active_all`	Percent all the CPUs spent in activity.	Gauge	N/A	Basic
`node_cpu_seconds_active_all_total`	Seconds all the CPUs spent in activity (excluding idle and wait).	Counter	N/A	Custom: `CPU time metrics`
`node_cpu_seconds_active_total`	Seconds each CPU spent in activity (excluding idle and wait).	Counter	`cpu`	Custom: `CPU time metrics`
`node_cpu_seconds_all_total`	Seconds for all CPUs usage.	Counter	`mode`	Custom: `CPU time metrics`
`node_cpu_seconds_total`	Seconds for each CPU's usage.	Counter	`cpu`, `mode`	Custom: `CPU time metrics`
`node_cpu_percent_active`	Percent each CPU spent in activity.	Gauge	`cpu`	Custom: `Per CPU` or `Detailed` metrics
`node_cpu_percent_all`	Percent CPU usage for all.	Gauge	`mode`,`user`	Custom: `Detailed` metrics
`node_cpu_percent`	Percent CPU usage for each CPU.	Gauge	`cpu`,`mode`,`user`	Custom: `Per CPU` or `Detailed` metrics

- The `Per CPU` option adds metrics with the `cpu` dimension.

- The `Detailed metrics` option adds metrics with `mode` dimension set to: `irq`, `softirq`, `steal`, `guest`, `guest_nice`, and `nice`.

Memory

With System Metrics enabled, Cribl Edge captures memory metrics including total, used, available, `swap_free`, and `swap_total`, with the option to toggle all memory states.

Metrics for memory include the following:

Name	Description	Type	Dimensions	Mode
`node_memory_MemTotal_bytes`	Memory information field `MemTotal_bytes`.	Gauge	N/A	Basic
`node_memory_Used_bytes`	Used memory in bytes.	Gauge	N/A	Basic
`node_memory_Used_percent`	Percent used memory.	Gauge	N/A	Basic
`node_memory_MemAvailable_bytes`	Memory information field `MemAvailable_bytes`.	Gauge	N/A	Basic
`node_memory_MemAvailable_percent`	Percent memory available.	Gauge	N/A	Basic
`node_memory_SwapFree_bytes`	Memory information field `SwapFree_bytes`.	Gauge	N/A	Basic
`node_memory_SwapTotal_bytes`	Memory information field `SwapTotal_bytes`.	Gauge	N/A	Basic
`node_vmstat_oom_kill`	`/proc/vmstat` information field `oom_kill`.	Gauge	N/A	Basic
`node_vmstat_pgfault`	`/proc/vmstat` information field `pgfault`.	Gauge	N/A	Basic
`node_vmstat_pgmajfault`	`/proc/vmstat` information field `pgmajfault`.	Gauge	N/A	Basic
`node_vmstat_pgpgin`	`/proc/vmstat` information field	Gauge	N/A	Basic

Name	Description	Type	Dimensions	Mode
	`pgpgin`.			
`node_vmstat_pgpgout`	`/proc/vmstat` information field `pgpgout`.	Gauge	N/A	Basic
`node_vmstat_pswpin`	`/proc/vmstat` information field `pswpin`.	Gauge	N/A	Basic
`node_vmstat_pswpout`	`/proc/vmstat` information field `pswpout`.	Gauge	N/A	Basic
`node_memory_Active_bytes`	Memory information field `Active_bytes`.	Gauge	N/A	Custom: `Detaile` metrics
`node_memory_Buffers_bytes`	Memory information field `Buffers_bytes`.	Gauge	N/A	Custom: `Detaile` metrics
`node_memory_Cached_bytes`	Memory information field `Cached_bytes`.	Gauge	N/A	Custom: `Detaile` metrics
`node_memory_MemFree_bytes`	Memory information field `MemFree_bytes`.	Gauge	N/A	Custom: `Detaile` metrics
`node_memory_SwapCached_bytes`	Memory information field `SwapCached_bytes`.	Gauge	N/A	Custom: `Detaile` metrics

Network

With System Metrics enabled, Cribl Edge captures bytes, packets, errors, and network connections over all interfaces. The **Custom** option allows you to filter interfaces, and to decide whether to select per-interface metrics and generate protocol metrics.

Metrics for networks include the following:

Name	Description	Type	Dimensions	M
`node_network_receive_bytes_all_total`	Network device statistic	Counter	N/A	E

Name	Description	Type	Dimensions	N
	`receive_bytes`.			
`node_network_receive_errs_all_total`	Network device statistic `receive_errs`.	Counter	N/A	E
`node_network_receive_packets_all_total`	Network device statistic `receive_packets`.	Counter	N/A	E
`node_network_transmit_bytes_all_total`	Network device statistic `transmit_bytes`.	Counter	N/A	E
`node_network_transmit_errs_all_total`	Network device statistic `transmit_errs`.	Counter	N/A	E
`node_network_transmit_packets_all_total`	Network device statistic `transmit_packets`.	Counter	N/A	E
`node_socket_tcp_established_total`	TCP established connections.	Counter	N/A	E
`node_network_receive_bytes_total`	Network device statistic `receive_bytes` per interface.	Counter	`device`	C F I
`node_network_receive_errs_total`	Network device statistic `receive_errs` per interface.	Counter	`device`	C F I
`node_network_receive_packets_total`	Network device statistic `receive_packets` per interface.	Counter	`device`	C F I
`node_network_transmit_bytes_total`	Network device statistic `transmit_bytes` per interface.	Counter	`device`	C F I
`node_network_transmit_errs_total`	Network device statistic `transmit_errs` per interface.	Counter	`device`	C F I

Name	Description	Type	Dimensions	N
node_network_transmit_packets_total	Network device statistic transmit_packets per interface.	Counter	device	C F I
node_network_receive_drop_all_total	Network device statistic receive_drop.	Counter	N/A	C D M
node_network_receive_drop_total	Network device statistic receive_drop per interface.	Counter	device	C D M
node_network_transmit_drop_all_total	Network device statistic transmit_drop.	Counter	N/A	C D M
node_network_transmit_drop_total	Network device statistic transmit_drop per interface.	Counter	device	C D M
node_socket_tcp_syn_sent_total	TCP sent packets total.	Counter	N/A	C D M
node_socket_tcp_syn_recv_total	TCP received packets total.	Counter	N/A	C D M
node_socket_tcp_fin_wait1_total	Total connections waiting for termination request from remote TCP.	Counter	N/A	C D M
node_socket_tcp_fin_wait2_total	Active TCP connections to be shut down.	Counter	N/A	C D M
node_socket_tcp_time_wait_total	Length of time to pass to be sure the remote TCP received the acknowledgement to terminate.	Counter	N/A	C D M
node_socket_tcp_close_total	Total TCP sockets with closed	Counter	N/A	C D

Name	Description	Type	Dimensions	N
	connections.			M
node_socket_tcp_last_ack_total	Total TCP sockets in state before the TCP connection is closed.	Counter	N/A	C D M
node_socket_tcp_listen_total	Total TCP sockets waiting for a connection request from any remote TCP/port.	Counter	N/A	C D M
node_socket_tcp_closing_total	Total TCP sockets waiting for connection termination request acknowledgement.	Counter	N/A	C D M
node_socket_tcp_none_total	Number of TCP sockets with no connections.	Counter	N/A	C D M
node_socket_udp_total	Number of UDP sockets in use.	Counter	N/A	C D M

Disk

With System Metrics enabled, Cribl Edge captures disk-used metrics – in percent, bytes read and written, and read and write operations – over all mounted disks. The **Custom** option allows you to filter devices, mountpoint, and filesystem type, and to decide whether to select per-device metrics and generate detailed metrics.

Metrics for Disk include the following:

Name	Description	Type	Dimension
node_disk_reads_completed_all_total	Total number of reads completed successfully.	Counter	N/A
node_disk_read_bytes_all_total	Total number of bytes read successfully.	Counter	N/A

Name	Description	Type	Dimension
node_disk_writes_completed_all_total	Network device statistic receive_packets.	Counter	N/A
node_disk_written_bytes_all_total	Total number of bytes written successfully.	Counter	N/A
node_filesystem_size_bytes_all	Filesystem size in bytes.	Gauge	N/A
node_filesystem_avail_bytes_all	Filesystem space available to non-root users in bytes.	Gauge	N/A
node_filesystem_used_bytes_all	Filesystem used space in bytes.	Gauge	N/A
node_filesystem_used_percent_all	Percent filesystem used space.	Gauge	N/A
node_filesystem_files_free_all	Filesystem free file nodes.	Gauge	N/A
node_filesystem_files_used_all	Filesystem total used file nodes.	Gauge	N/A
node_filesystem_files_used_percent_all	Percent Filesystem used in all disks.	Gauge	N/A
node_filesystem_size_bytes_all	Filesystem size in bytes.	Gauge	N/A
node_filesystem_used_bytes	Filesystem used space in bytes per disk.	Gauge	device, fstype, mountpoin
node_filesystem_used_percent	Percent Filesystem used per disk.	Gauge	device, fstype, mountpoin
node_disk_reads_completed_total	Total number of reads completed successfully per disk.	Counter	device

Name	Description	Type	Dimension
node_disk_read_bytes_total	Total number of bytes read successfully per disk.	Counter	device, fstype, mountpoin
node_disk_writes_completed_total	Network device statistic receive_packets per disk.	Counter	device, fstype, mountpoin
node_disk_written_bytes_total	Total number of bytes written successfully per disk.	Counter	device, fstype, mountpoin
node_disk_discards_completed_all_total	Total number of discards completed successfully.	Counter	N/A
node_disk_discards_completed_total	Total number of discards completed successfully per disk.	Counter	device, fstype, mountpoin
node_disk_discards_merged_all_total	Total number of discards merged.	Counter	N/A
node_disk_read_time_seconds_total	Total number of seconds spent by all reads per disk.	Counter	device, fstype, mountpoin
node_disk_write_time_seconds_all_total	Total number of seconds spent by all writes.	Counter	N/A
node_disk_read_time_seconds_all_total	Total number of seconds spent by all reads.	Counter	N/A
node_disk_write_time_seconds_total	Total number of seconds spent by all writes per disk.	Counter	device, fstype, mountpoin
node_disk_reads_merged_all_total	Total number of reads merged.	Counter	N/A

Name	Description	Type	Dimension
node_disk_writes_merged_all_total	Total number of writes merged.	Counter	N/A
node_disk_reads_merged_total	Total number of reads merged per disk.	Counter	device, fstype, mountpoin
node_disk_writes_merged_total	Total number of writes merged per disk.	Counter	device, fstype, mountpoin
node_disk_discards_merged_total	Total number of discards merged per disk.	Counter	device, fstype, mountpoin
node_disk_io_time_seconds_all_total	Total seconds spent doing I/Os.	Counter	N/A
node_disk_io_time_seconds_total	Total seconds spent doing I/Os per disk.	Counter	device, fstype, mountpoin
node_disk_io_now_all	The number of I/Os currently in progress.	Gauge	N/A
node_disk_io_now	The number of I/Os currently in progress per disk.	Counter	device, fstype, mountpoin
node_disk_io_time_weighted_seconds_all_total	Weighted number of seconds spent doing I/Os per device.	Counter	device, fstype, mountpoin
node_filesystem_size_bytes	Filesystem size in bytes per disk.	Gauge	device, fstype, mountpoin

Name	Description	Type	Dimension
`node_filesystem_avail_bytes`	Filesystem space available to non-root users in bytes per disk.	Gauge	`device`, `fstype`, `mountpoin`
`node_filesystem_files_free`	Filesystem free file nodes per disk.	Gauge	`device`, `fstype`, `mountpoin`
`node_filesystem_files_used_percent`	Percent Filesystem used per disk.	Gauge	`device`, `fstype`, `mountpoin`
`node_filesystem_size_bytes`	Filesystem size in bytes per disk.	Gauge	`device`, `fstype`, `mountpoin`

Container

With System Metrics enabled, Cribl Edge generates Docker information with CPU, memory, network, and disk metrics for running containers. Optionally, you can customize which containers to generate metrics from.

Metrics for Container include the following:

Name	Description	Type
`container_start_time_seconds`	UNIX time (seconds since the epoch) when the container was started.	Counte
`container_finish_time_seconds`	UNIX time (seconds since the epoch) when the container was stopped. Only for non-running containers.	Counte
`container_fs_reads_bytes_all_total`	Total bytes read for all disk devices.	Counte
`container_memory_usage_percent`	Percent of available memory being used.	Gauge
`container_network_receive_bytes_all_total`	Total bytes received for all network interfaces.	Counte

Name	Description	Type
container_network_receive_errors_all_total	Total number of errors received for all network interfaces.	Counte
container_network_receive_packets_all_total	Total number of packets received for all network interfaces.	Counte
container_network_transmit_bytes_all_total	Total bytes transmitted for all network interfaces.	Counte
container_network_transmit_errors_all_total	Total number of errors transmitted for all network interfaces.	Counte
container_network_transmit_packets_all_total	Total number of packets transmitted for all network interfaces.	Counte
container_memory_total_bytes	Total number of memory bytes available.	Counte
container_cpu_user_seconds_total	Number of seconds the container has been on the CPU running user code.	Counte
container_cpu_system_seconds_total	Number of seconds the container has been on the CPU running kernel code.	Counte
container_fs_reads_bytes_total	Total bytes read per device	Counte
container_fs_writes_bytes_all_total	Total bytes written for all disk devices	Counte
container_fs_writes_bytes_total	Total bytes written per device	Counte

Name	Description	Type
`container_fs_reads_all_total`	Total number of read operations for all disk devices.	Counte
`container_fs_writes_total`	Total number write operations per device.	Counte
`container_memory_mapped_file`	Total bytes writted for all disk devices.	Counte
`container_memory_max_usage_bytes`	Highest seen value of the `container_memory_usage_bytes` metric.	Counte
`container_memory_pgin`	Total number of memory page-in events.	Counte
`container_mem.pgpgout`	Total number of memory pages paged out.	Counte
`container_memory_pgfault`	Total number of major page faults.	Counte
`container_memory_pgmajfault`	Total number of minor page faults.	Counte
`container_memory_usage_bytes`	Number of memory bytes used.	Counte
`container_network_receive_dropped_all_total`	Total number of receives dropped for all network interfaces.	Counte

Name	Description	Type
`container_network_transmit_dropped_all_total`	Total number of transmits dropped for all network interfaces.	Counte

Process Metrics

With Process Metrics enabled, Cribl Edge captures process-specific metrics from Linux servers and reports them as events. This allows you to monitor specific processes on Cribl.Cloud instances. You can generate events for any process object.

For information on how to configure the System Metrics Source to generate process-specific metrics, check out the Process Metrics section of the System Metrics page.

Process-specific metrics are **not** affected by the **Host Metrics** detail setting.

Process-specific metrics include the following:

Name	Description	Type	Dimensions
`process_num_threads`	The number of threads.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_open_filedesc`	The number of file descriptors.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_write_bytes`	The number of bytes which this process caused to be sent to the storage layer.	Counter	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_read_bytes`	The number of bytes this process actually fetched from the storage layer. This number is	Counter	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`

Name	Description	Type	Dimensions
	accurate for block-backed filesystems.		
process_major_page_faults	The number of major faults for this process that required loading a memory page from disk.	Counter	process_cmdline, process_set, process_uid, process_gid, process_service
process_minor_page_faults	The number of minor faults for this process that have not required loading a memory page from disk.	Counter	process_cmdline, process_set, process_uid, process_gid, process_service
process_voluntary_context_switches	The number of voluntary context switches.	Counter	process_cmdline, process_set, process_uid, process_gid, process_service
process_nonvoluntary_context_switches	The number of involuntary context switches.	Counter	process_cmdline, process_set, process_uid, process_gid, process_service
process_cpu_usage	The process's CPU usage, expressed as a percentage of total CPU power.	Gauge	process_cmdline, process_set, process_uid, process_gid, process_service
process_cpu_seconds	The process's CPU usage, based on user time and system time.	Counter	process_cmdline, process_set, process_uid, process_gid, process_service
process_resident_memory_bytes	The amount of memory used, in bytes. Includes the pages that count toward text, data, or stack space.	Gauge	process_cmdline, process_set, process_uid, process_gid, process_service

Name	Description	Type	Dimensions
	Does not include pages that haven't been demand-loaded in or are swapped out.		
`process_virtual_memory_bytes`	The process's virtual memory size.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_swapped_memory_bytes`	The process's swapped memory size.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_memory_bytes`	The total amount of memory used by the process, in bytes.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_memory_usage`	The total amount of memory used by the process, as a percentage of total memory.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`
`process_start_time`	The time that the process started, derived by adding the start time to the boot time, making it relative to epoch.	Gauge	`process_cmdline`, `process_set`, `process_uid`, `process_gid`, `process_service`

16.8. Windows System Metrics Details

Events generated by the Windows Metrics Source include metrics metadata to designate dimension and metric fields. The `host` field contains the hostname, and is included as a dimension in all of them. The collectors include:

- System
- CPU
- Memory
- Network
- Disk
- Process

In the Source's configuration modal, You can set the level of detail for each type of metrics:

- **Basic** enables minimal metrics, averaged or aggregated.
- **All** enables full, detailed metrics, specified for individual CPUs, interfaces, and so on.
- **Custom** displays sub-menus and buttons from which you can choose a level of detail (Basic, All, Custom, or Disabled) for each type of event.
- **Disabled** means that no metrics will be generated.

Basic and **Custom** have different meanings depending on event type and will be described under each section below.

The tables outline the metrics emitted for each mode (**Basic** or **Custom**) and where applicable, the dimensions (to indicate where the metrics are coming from).

System

With System Metrics enabled, Cribl Edge captures CPU load averages, uptime, and count. The **Custom** option option allows you to include detailed metrics. These are Windows-specific metrics including OS information, system uptime, CPU architecture, etc.

Metrics for the overall system include the following:

Name	Description	Type	Dimensions	M
`windows_cs_logical_processors`	Number of installed	Gauge	N/A	Ba

Name	Description	Type	Dimensions	Me
	logical processors.			
windows_cs_physical_memory_bytes	Total installed physical memory.	Gauge	N/A	Ba
windows_os_info	Contains full product name & version in labels.	Gauge	product, version	Ba
windows_os_physical_memory_free_bytes	Bytes of physical memory currently unused and available.	Gauge	N/A	Ba
windows_os_processes	Number of process contexts currently loaded or running on the operating system.	Gauge	N/A	Ba
windows_system_processor_queue_length	Number of threads in the processor queue.	Gauge	N/A	Ba
windows_system_threads	Number of Windows system threads.	Gauge	N/A	Ba
windows_cs_hostname	Labeled system hostname information.	Gauge	hostname, domain, fqdn	Cu De
windows_cpu_info	Labeled CPU information.	Gauge	architecture, device_id, description, family,	Cu De

Name	Description	Type	Dimensions	M
			`l2_cache_size`, `l3_cache_size`, `name`	
`windows_os_paging_limit_bytes`	Total number of bytes that can be stored in the operating system paging files.	Gauge	N/A	Cu De
`windows_os_paging_free_bytes`	Number of bytes that can be mapped into the operating system paging files without causing any other pages to be swapped out.	Gauge	N/A	Cu De
`windows_os_processes_limit`	Maximum number of process contexts the operating system can support.	Gauge	N/A	Cu De
`windows_os_process_memory_limit_bytes`	Maximum number of bytes of memory that can be allocated to a process.	Gauge	N/A	Cu De
`windows_os_virtual_memory_bytes`	Bytes of virtual memory.	Gauge	N/A	Cu De
`windows_system_exception_dispatches_total`	Total exceptions	Counter	N/A	Cu De

Name	Description	Type	Dimensions	Mo
	dispatched by the system.			
`windows_system_system_calls_total`	Total combined calls to Windows NT system service routines by all processes running on the computer.	Counter	N/A	Cu De
`windows_system_system_up_time`	Time of last boot of system.	Gauge	N/A	Cu De

CPU

Basic level captures active, user, system, idle, and `iowait` percentages over all CPUs.

Custom level toggles the following on or off: **Per CPU metrics**, **Detailed metrics** (i.e., metrics for all CPU states), and **CPU time metrics** (i.e., raw, monotonic CPU time counters).

Metrics for CPUs include the following:

Name	Description	Type	Dimensions	Mo
`windows_cpu_percent_active_all`	CPU percent active usage	Gauge	`core`, `mode`	Ba
`windows_cpu_percent_active`	CPU percent active usage per CPU	Gauge	`core`, `mode`	Ba Cu Pe an ti me
`windows_cpu_percent`	CPU percent active usage	Gauge	`core`, `mode` is set to `user`, `idle`, `privileged`, `interrupt`, `dpc`	Ba Cu Pe an ti me

Name	Description	Type	Dimensions	Mo
windows_cpu_parking_status	Parking Status represents whether a processor is parked or not.	Counter	core	Ba Cu Pe an ti me
windows_cpu_core_frequency_mhz	Core frequency in megahertz.	Gauge	core	Ba Cu Pe an ti me
windows_cpu_time_all_total	Sum of all cpu_time across all cores.	Gauge	mode	Ba Cu CP me
windows_cpu_cstate_seconds_total	Time spent in low-power idle state.	Counter	core, state	Cu Pe an De me
windows_cpu_time_total	Time that processor spent in different modes (idle, user, system etc.).	Counter	core, mode	Cu Pe an ti me
windows_cpu_interrupts_total	Total number of received and serviced hardware interrupts.	Counter	core	Cu Pe or De me
windows_cpu_dpcs_total	Total number of received and serviced deferred procedure calls (DPCs).	Counter	core	Cu Pe or De me

Name	Description	Type	Dimensions	Mo
windows_cpu_clock_interrupts_total	Total number of received and serviced clock tick interrupts.	Counter	core	Cu Pe or De me
windows_cpu_idle_break_events_total	Total number of time processor was woken from idle.	Counter	core	Cu Pe or De me
windows_cpu_processor_performance	Average performance of the processor while it is executing instructions.	Gauge.	core	Cu Pe an De me
windows_cpu_percent_processor_performance	Average performance of the processor while it is executing instructions, as a percentage of the nominal performance of the processor.	Gauge.	core	Cu Pe an De me
windows_cpu_percent_processor_utility	Amount of work a processor is completing, as a percentage of the amount of work the processor could complete if it were running at	Gauge.	core	Cu Pe an De me

Name	Description	Type	Dimensions	Mo
	its nominal performance and never idle.			
windows_cpu_average_idle_time	Processor idle time.	Gauge	mode	Cu Pe an De me
windows_cpu_percent_privilege_utility	Amount of work a processor is completing while executing in privileged mode.	Gauge	mode	Cu Pe an De me
windows_cpu_interrupts_total_per_sec	Total number of received and serviced hardware interrupts, computed average on a per second interval.	Gauge	mode	Cu Pe an De me
windows_cpu_dpcs_total_per_sec	Total number of received and serviced deferred procedure calls (DPCs), computed average on a per second interval.	Gauge	mode	Cu Pe an De me
windows_cpu_clock_interrupts_total_per_sec	Total number of received and serviced clock tick interrupts, computed average on a	Gauge	mode	Cu Pe an De me

Name	Description	Type	Dimensions	Mo
	per second interval.			
`windows_cpu_idle_break_events_total_per_sec`	Total number of time processor was woken from idle, computed average on a per second interval.	Gauge	`mode`	Cu Pe an De me
`windows_cpu_percent_all_total`	Core frequency in megahertz.	Gauge	`mode` is set to `dpc`, `idle`, `interrupt`, `privilege`, `user`,`active`	Cu CP me

Memory

Basic level captures captures total, used, available, `swap_free`, and `swap_total`.

Custom level toggles **Detailed metrics** on or off. (These are metrics for all memory states.)

Metrics for memory include the following:

Name	Description	Type	Dimen
`windows_memory_available_bytes`	Physical memory that is immediately available for allocation to a process or for system use. This is the sum of the standby (cached), free, and zero page lists.	Gauge	N/A
`windows_memory_cache_bytes`	Number of bytes currently being used by the filesystem cache	Gauge	N/A

Name	Description	Type	Dimen
windows_memory_cache_bytes_peak	Maximum number of CacheBytes after the system was last restarted.	Gauge	N/A
windows_memory_cache_faults_total	Faults that occur when a page sought in the filesystem cache is not found there and must be retrieved elsewhere in memory (soft fault) or from disk (hard fault).	Counter	N/A
windows_memory_commit_limit	Bytes of virtual memory that can be committed without having to extend paging files.	Gauge	N/A
windows_memory_committed_bytes	Bytes of committed virtual memory.	Gauge	N/A
windows_memory_pool_paged_allocs_total	Calls to allocate space in the paged pool, regardless of the amount of space allocated in each call.	Counter	N/A
windows_memory_pool_paged_bytes	Number of bytes in the paged pool.	Gauge	N/A
windows_memory_pool_paged_resident_bytes	The size, in bytes, of the portion of the paged pool that is currently resident and active in physical	Gauge	N/A

Name	Description	Type	Dimen
	memory. The paged pool is an area of the system virtual memory used for objects that can be written to disk when they are not being used.		
windows_memory_demand_zero_faults_total	Number of Zeroed pages required to satisfy faults. Windows uses zeroed pages as a security measure to prevent processes from seeing data stored by earlier processes that previously used the memory space.	Counter	N/A
windows_memory_free_and_zero_page_list_bytes	Physical memory allocated to free and zero pages, in bytes. This memory does not contain cached data. It is immediately available for allocation to a process or system use.	Gauge	N/A
windows_memory_free_system_page_table_entries	Page table entries not being used by the system.	Gauge	N/A
windows_memory_modified_page_list_bytes	Physical memory, in bytes, assigned to the modified page list. This	Gauge	N/A

Name	Description	Type	Dimen
	memory contains cached data and code that is not actively in use by processes, the system, and the system cache. This memory needs to be written out before it will be available for allocation to a process or for system use.		
windows_memory_page_faults_total	Overall rate at which faulted pages are handled by the processor.	Counter	N/A
windows_memory_swap_page_reads_total	Disk page reads (a single read operation reading several pages is still only counted once).	Counter	N/A
windows_memory_swap_pages_read_total	Pages read across all page reads (i.e., counting all pages read even if they are read in a single operation).	Counter	N/A
windows_memory_swap_pages_written_total	Pages written across all page writes (i.e., counting all pages written even if they are written in a single operation).	Counter	N/A
windows_memory_swap_page_operations_total	Total number of swap page read	Counter	N/A

Name	Description	Type	Dimen
	and writes (PagesPersec).		
windows_memory_swap_page_writes_total	Disk page writes (a single write operation writing several pages is still only counted once).	Counter	N/A
windows_memory_pool_nonpaged_allocs_total	The number of calls to allocate space in the non-paged pool. The non-paged pool is an area of system memory area for objects that cannot be written to disk, and must remain in physical memory as long as they are allocated	Counter.	N/A
windows_memory_pool_nonpaged_bytes	Non-paged pool, in bytes. The non-paged pool is an area of the system virtual memory that is used for objects that cannot be written to disk, but must remain in physical memory as long as they are allocated.	Gauge	N/A
windows_memory_standby_cache_core_bytes	Physical memory, in bytes, that is assigned to the core standby cache page lists. This memory contains cached data and code	Gauge	N/A

Name	Description	Type	Dimen
	that is not actively in use by processes, the system, and the system cache. It is immediately available for allocation to a process or for system use. If the system runs out of available free and zero memory, memory on lower priority standby cache page lists will be repurposed before memory on higher priority standby cache page lists.		
windows_memory_standby_cache_normal_priority_bytes	Physical memory, in bytes, that is assigned to the normal priority standby cache page lists. This memory contains cached data and code that is not actively in use by processes, the system, and the system cache. It is immediately available for allocation to a process or for system use. If the system runs out of available free and zero memory, memory on lower priority standby cache page lists will be	Gauge	N/A

Name	Description	Type	Dimen
	repurposed before memory on higher priority standby cache page lists.		
windows_memory_standby_cache_reserve_bytes	Pysical memory, in bytes, that is assigned to the reserve standby cache page lists. This memory contains cached data and code that is not actively in use by processes, the system and the system cache. It is immediately available for allocation to a process or for system use. If the system runs out of available free and zero memory, memory on lower priority standby cache page lists will be repurposed before memory on higher priority standby cache page lists.	Gauge	N/A
windows_memory_system_cache_resident_bytes	The size, in bytes, of the portion of the system file cache which is currently resident and active in physical memory.	Gauge	N/A
windows_memory_system_code_resident_bytes	The size, in bytes, of the pageable operating	Gauge	N/A

Name	Description	Type	Dimen
	system code that is currently resident and active in physical memory. This value is a component of Memory/System Code Total Bytes. Memory/System Code Resident Bytes (and Memory/System Code Total Bytes) does not include code that must remain in physical memory and cannot be written to disk.		
`windows_memory_system_code_total_bytes`	The size, in bytes, of the pageable operating system code currently mapped into the system virtual address space. This value is calculated by summing the bytes in Ntoskrnl.exe, Hal.dll, the boot drivers, and filesystems loaded by Ntldr/osloader. This counter does not include code that must remain in physical memory and cannot be written to disk.	Gauge	N/A
`windows_memory_system_driver_resident_bytes`	The size, in bytes, of the	Gauge	N/A

Name	Description	Type	Dimen
	physical memory being used by device drivers. It is the working set (physical memory area) of the drivers. This value is a component of Memory/System Driver Total Bytes, which also includes driver memory that has been written to disk. Neither Memory/System Driver Resident Bytes nor Memory/System Driver Total Bytes includes memory that cannot be written to disk.		
`windows_memory_system_driver_total_bytes`	The size, in bytes, of the pageable virtual memory currently being used by device drivers. Pageable memory can be written to disk when it is not being used. It includes both physical memory (Memory/System Driver Resident Bytes) and code and data paged to disk. It is a component of Memory/System Code Total Bytes.	Gauge	N/A

Name	Description	Type	Dimen
windows_memory_transition_faults_total	Rate at which page faults are resolved, by recovering pages that were being used by another process sharing the page, or were on the modified page list or the standby list, or were being written to disk at the time of the page fault. The pages were recovered without additional disk activity. Transition faults are counted in numbers of faults; because only one page is faulted in each operation, it is also equal to the number of pages faulted.	Counter	N/A
windows_memory_transition_pages_repurposed_total	Rate at which the number of transition cache pages were reused for a different purpose. These pages would have otherwise remained in the page cache to provide a (fast) soft fault (instead of retrieving it from backing store) in the event the page was	Counter	N/A

Name	Description	Type	Dimen
	accessed in the future.		
windows_memory_write_copies_total	The number of page faults caused by attempting to write that were satisfied by copying the page from elsewhere in physical memory.	Counter	N/A
windows_memory_used_percent	Percent memory used.	Gauge	N/A
windows_memory_available_percent	Percent memory available.	Gauge	N/A
windows_memory_cache_faults_per_sec	Rate of cache faults computed per sec.	Gauge	N/A
windows_memory_demand_zero_faults_per_sec	Rate of Zeroed pages required to satisfy faults computed per sec.	Gauge	N/A
windows_memory_page_faults_per_sec	Rate at which faulted pages are handled by the processor computed per sec.	Gauge	N/A
windows_memory_page_reads_per_sec	Disk page reads computed per sec.	Gauge	N/A
windows_memory_pages_input_per_sec	Disk page reads computed per sec.	Gauge	N/A
windows_memory_pages_output_per_sec	Pages written across all page writes computed per sec.	Gauge	N/A

Name	Description	Type	Dimen
windows_memory_pages_per_sec	Total number of swap page read and writes computed per sec.	Gauge	N/A
windows_memory_page_writes_per_sec	Disk page writes computed per sec.	Gauge	N/A
windows_memory_transition_faults_sec	Rate at which page faults are resolved computed per sec.	Gauge	N/A
windows_memory_transition_pages_repurposed_per_sec	Rate at which the number of transition cache pages were reused for a different purpose computed per sec.	Gauge	N/A
windows_memory_write_copies_per_sec	Rate at which the the number of page faults caused by attempting to write that were satisfied by copying the page from elsewhere in physical memory computed per sec.	Gauge	N/A

Network

Basic level captures bytes, packets, errors, and connections over all interfaces.

Custom level exposes the following:

- The **Interface filter**, which specifies which network interfaces to include or exclude. (An empty filter will include all metrics.)
- **Per interface metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off. If on, the **Protocol metrics** toggle appears, allowing you to choose whether to generate metrics for ICMP, ICMPMsg, IP, TCP, UDP, and UDPLite.

Metrics for networks include the following:

Name	Description	Type	Dim
`windows_net_packets_outbound_discarded_total`	Total outbound packets to be discarded even though no errors had been detected to prevent transmission.	Counter	`nic`
`windows_net_packets_outbound_errors_total`	Total packets that could not be transmitted due to errors.	Counter	`nic`
`windows_net_packets_received_discarded_total`	Total inbound packets that were chosen to be discarded even though no errors had been detected to prevent delivery.	Counter	`nic`
`windows_net_packets_received_errors_total`	Total packets that could not be received due to errors.	Counter	`nic`
`windows_net_packets_received_unknown_total`	Total packets received by interface that were	Counter	`nic`

Name	Description	Type	Dim
	discarded because of an unknown or unsupported protocol.		
windows_net_packets_received_non_unicast_total	Total non-unicast (subnet broadcast or subnet multicast) packets that are delivered to a higher-layer protocol.	Counter	nic
windows_net_packets_received_unicast_total	Total subnet-unicast packets that are delivered to a higher-layer protocol.	Counter	nic
windows_net_packets_sent_unicast_total	Total packets requested to be transmitted to subnet-unicast addresses by higher-level protocols.	Counter	nic
windows_net_packets_sent_non_unicast_total_per_sec	Total packets that are requested to be transmitted to nonunicast (subnet broadcast or subnet multicast) addresses by higher-level protocols.	Gauge	nic

Name	Description	Type	Dim
windows_net_bytes_received_total	Total bytes received by interface.	Counter	nic
wwindows_net_bytes_received_total_per_sec	Total bytes received by interface computed per sec.	Gauge	nic
windows_net_bytes_sent_total	Total bytes transmitted by interface.	Counter	nic
windows_net_bytes_sent_total_per_sec	Total bytes transmitted by interface computed per sec.	Gauge	nic
windows_net_bytes_total	Total bytes received and transmitted by interface.	Counter	nic
windows_net_bytes_total_per_sec	Total bytes received and transmitted by interface per sec.	Gauge	nic
windows_net_packets_received_total_per_sec	Total packets received by interface computed per sec.	Counter	nic
windows_net_bytes_total_per_sec	Total bytes received and transmitted by interface per sec.	Gauge	nic

Name	Description	Type	Dim
windows_net_packets_received_non_unicast_total_per_sec	Rate at which non-unicast (subnet broadcast or subnet multicast) packets are delivered to a higher-layer protocol computed per sec.	Gauge	nic
windows_net_packets_total	Total packets received and transmitted by interface.	Counter	nic
windows_net_packets_total_per_sec	Total packets received and transmitted by interface computed per sec.	Gauge	nic
windows_net_packets_sent_total	Total packets transmitted by interface.	Counter	nic
windows_net_packets_sent_total_per_sec	Total packets transmitted by interface computed per sec.	Gauge	nic
windows_net_packets_sent_unicast_total_per_sec	Rate at which packets are requested to be transmitted to subnet-unicast addresses by higher-level protocols computed per sec.	Gauge	nic

Name	Description	Type	Dim
windows_net_packets_sent_non_unicast_total	Rate at which packets that are requested to be transmitted to nonunicast (subnet broadcast or subnet multicast) addresses by higher-level protocols per sec.	Counter	nic
windows_net_current_bandwidth_bytes	Estimate of the interface's current bandwidth in bytes per second.	Gauge	nic
windows_tcp_connection_failures_all_total	Number of times TCP connections have made a direct transition to the CLOSED state from the SYN-SENT state or the SYN-RCVD state, plus the number of times TCP connections have made a direct transition from the SYN-RCVD state to the LISTEN state.	Counter	af
windows_tcp_connections_active_all_total	Number of times TCP connections have made a	Counter	af

Name	Description	Type	Dim
	direct transition from the CLOSED state to the SYN-SENT state.		
windows_tcp_connections_established	Number of TCP connections for which the current state is either ESTABLISHED or CLOSE-WAIT.	Gauge	af
windows_tcp_connections_passive_all_total	Number of times TCP connections have made a direct transition from the LISTEN state to the SYN-RCVD state.	Counter	af
windows_tcp_connections_reset_total	Number of times TCP connections have made a direct transition from the LISTEN state to the SYN-RCVD state.	Counter	af
windows_tcp_segments_total	Total segments sent or received using the TCP protocol.	Counter	af
windows_tcp_segments_received_all_total	Total segments received	Counter	af

Name	Description	Type	Dim
	using the TCP protocol.		
windows_tcp_segments_retransmitted_all_total	Total segments retransmitted using the TCP protocol.	Counter	af
windows_tcp_segments_sent_all_total	Total segments sent using the TCP protocol.	Counter	af
windows_tcp_segments_all_total_per_sec	Total segments sent or received using the TCP protocol computed per sec.	Gauge	af
windows_tcp_segments_received_all_total_per_sec	Total segments received using the TCP protocol computed per sec.	Gauge	af
windows_tcp_segments_retransmitted_all_total_per_sec	Total segments retransmitted using the TCP protocol computed per sec.	Gauge	af

Name	Description	Type	Dim
windows_tcp_segments_sent_all_total_per_sec	Total segments sent using the TCP protocol computed per sec.	Gauge	af
windows_net_datagrams_all_total	Total datagrams sent or received using the UDP protocol.	Counter	af
windows_net_datagrams_no_port_all_total	Rate of received UDP datagrams for which there was no application at the destination port.	Gauge	af
windows_net_datagrams_received_all_total	Rate at which UDP datagrams are delivered to UDP users.	Gauge	af
windows_net_datagrams_received_errors_all_total	Number of received UDP datagrams that could not be delivered excluding errors due to lack of an application at	Gauge	af

Name	Description	Type	Dim
	the destination port.		
windows_net_datagrams_sent_all_total	Total UDP datagrams sent from the entity.	Gauge	af
windows_net_datagrams_no_port_all_total_per_sec	Rate of received UDP datagrams for which there was no application at the destination port computed per sec	Gauge	af
windows_net_datagrams_received_all_total_per_sec	Rate at which UDP datagrams are delivered to UDP users computed per sec	Gauge	af
windows_net_datagrams_received_errors_all_total_per_sec	Rate at which received UDP datagrams that could not be delivered, excluding errors due to lack of an application at the destination port, computed per sec.	Gauge	af

Name	Description	Type	Dim
windows_net_datagrams_sent_all_total_per_sec	Rate at which UDP datagrams sent from the entity computed per sec.	Gauge	af

Disk

Basic level captures disk usage (%), bytes read and written, and read and write operations, over all mounted disks.

Custom level exposes the following:

- The **Volume filter**, specifying which Windows volumes to include or exclude. Supports wildcards and ! (not) operators. An empty filter will include all volumes.
- **Per volume metrics**, which toggle on or off.
- **Detailed metrics**, which toggle on or off.

Metrics for Disk include the following:

Name	Description	Type	Dimens
windows_logical_disk_requests_queued	Outstanding requests on the disk at the time the performance data is collected	Gauge	volume
windows_logical_disk_read_bytes_total	Rate at which bytes are transferred from the disk during read operations.	Counter	volume
windows_logical_disk_reads_total	Rate of read operations	Counter	volume

Name	Description	Type	Dimens
	on the disk.		
windows_logical_disk_write_bytes_total	Rate at which bytes are transferred to the disk during write operations.	Counter	volume
windows_logical_disk_writes_total	Rate of write operations on the disk.	Counter	volume
windows_logical_disk_write_latency_seconds_total	Shows the average time, in seconds, of a write operation to the disk.	Counter	volume
windows_logical_disk_read_latency_seconds_total	Shows the average time, in seconds, of a read operation from the disk.	Counter	volume
windows_logical_disk_read_write_latency_seconds_total	Shows the time, in seconds, of the average disk transfer.	Counter	volume
windows_logical_disk_read_seconds_total	Seconds the disk was busy servicing read request.	Counter	volume
windows_logical_disk_idle_seconds_total	Seconds the disk was idle (not servicing	Counter	volume

Name	Description	Type	Dimens
	read/write requests).		
windows_logical_disk_split_ios_total	Number of I/Os to the disk split into multiple I/Os.	Counter	volume
windows_logical_disk_percent_read_time	Percent rate of read operations on the disk.	Gauge	volume
windows_logical_disk_percent_write_time	Percent write operations on the disk.	Gauge	volume
windows_logical_disk_percent_time	Percent time the disk was in read + write operations	Gauge	volume
windows_logical_disk_percent_time	Percent time the disk was idle.	Gauge	volume
windows_logical_disk_percent_free_space	Percent space free on volume.	Gauge	volume
windows_logical_disk_average_disk_sec_per_transfer	Measures the average time of data reads and writes on the disk.	Gauge	volume
windows_logical_disk_average_disk_sec_per_read	Measures the average rate of disk read requests that are executed per second on a specific	Gauge	volume

Name	Description	Type	Dimens
	physical disk.		
windows_logical_disk_average_disk_sec_per_write	Indicates how fast data is being written on average for a specific logical disk.	Gauge	volume
windows_logical_disk_split_ios_per_sec	Rate the I/Os to the disk were split into multiple I/Os per sec.	Gauge	volume
windows_logical_disk_bytes_per_sec	Exposes the rate bytes are transferred to or from the disk during write or read operations per sec.	Gauge	volume
windows_logical_disk_read_bytes_per_sec	Exposes the rate bytes are transferred to or from the disk during read operations per sec.	Gauge	volume
windows_logical_disk_reads_per_sec	Exposes the rate of read operations on the disk per sec.	Gauge	volume
windows_logical_disk_transfers_per_sec	How fast data is being read and written for a specific	Gauge	volume

Name	Description	Type	Dimens
	logical disk per sec.		
`windows_logical_disk_write_bytes_per_se`	Exposes the rate at which bytes are transferred from the disk during write operations per sec.	Gauge	`volume`

The `windows_logical_disk_free_bytes` and `windows_logical_disk_size_bytes` metrics are not updated in real time and might have a delay of 10-15min. This is the same behavior as the Windows performance counters.

Process Metrics

With Process Metrics enabled, Cribl Edge captures process-specific metrics from Windows servers and reports them as events. This allows you to monitor specific processes on Cribl.Cloud instances. You can generate events for any process object.

Process-specific metrics are **not** affected by the **Host Metrics** detail setting.

For information on how to configure the Windows Metrics Source to generate process-specific metrics, check out the section of the Windows Metrics page.

Process-specific metrics include the following:

Name	Description	Type	Dimensions
`process_start_time`	Time the process started.	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_cpu_time_total`	Elapsed time that the process's threads have spent executing instructions in either privileged	counter	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`

Name	Description	Type	Dimensions
	mode or user mode. Included in this count is code executed to handle some hardware interrupts and trap conditions.		
`process_handles`	Total number of handles the process has open. This number is the sum of the handles currently open by each thread in the process.	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_io_bytes_total`	Total number of bytes issued to I/O operations in either `read`, `write`, or `other` mode. This property counts all I/O activity generated by the process to include file, network, and device I/Os. `Read` and `write` modes include data operations; `other` mode includes those that don't involve data, like control operations.	counter	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_io_operations_total`	Total number of I/O operations issued in either `read`, `write`, or `other` mode. This property counts all I/O activity generated by the process to include file, network, and device I/Os. Read and write mode includes data	counter	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`

Name	Description	Type	Dimensions
	operations; other mode includes those that do not involve data, such as control operations.		
process_page_faults_total	Total number of page faults by threads executing in this process. A page fault occurs when a thread refers to a virtual memory page that is not in its working set in main memory. This can cause the page not to be fetched from disk if it is on the standby list and hence already in main memory, or if it is in use by another process with which the page is shared.	counter	process_set, process_ppid, process_pid, process_cmdline, process_exe_path, process_service_name
process_page_file_bytes	Current number of bytes this process has used in the paging files. Paging files are used to store pages of memory used by the process that are not contained in other files. Paging files are shared by all processes, and lack of space in paging files can prevent other processes from allocating memory.	gauge	process_set, process_ppid, process_pid, process_cmdline, process_exe_path, process_service_name

Name	Description	Type	Dimensions
process_pool_bytes	Last observed number of bytes in the paged or nonpaged pool. The paged pool is an area of system memory (physical memory used by the operating system) for objects that can be written to disk when they are not being used. The nonpaged pool is an area of system memory (physical memory used by the operating system) for objects that cannot be written to disk, but must remain in physical memory as long as they are allocated. Nonpaged pool bytes are calculated differently than paged pool bytes, so they may not equal the total of paged pool bytes.	gauge	process_set, process_ppid, process_pid, process_cmdline, process_exe_path, process_service_name
process_priority_base	Current base priority of this process. Threads within a process can raise and lower their own base priority, relative to the process's base priority.	gauge	process_set, process_ppid, process_pid, process_cmdline, process_exe_path, process_service_name
process_private_bytes	Current number of bytes this process has allocated that	gauge	process_set, process_ppid, process_pid, process_cmdline,

Name	Description	Type	Dimensions
	can't be shared with other processes.		`process_exe_path`, `process_service_name`
`process_threads`	Number of threads currently active in this process. Every running process has at least one thread.	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_virtual_bytes`	Current size, in bytes, of the virtual address space that the process is using. Use of virtual space doesn't imply use of either disk or main memory pages. Virtual space is finite and when the process uses too much, it can limit its ability to load libraries.	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_working_set_private_byte`	Size of the working set, in bytes, that is used for this process only and not shared or shareable by other processes.	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`
`process_working_set_peak_bytes`	Maximum size, in bytes, of the working set of this process at any point in time. The working set is the set of memory pages touched recently by the threads in the process. If free memory is above a threshold, pages	gauge	`process_set`, `process_ppid`, `process_pid`, `process_cmdline`, `process_exe_path`, `process_service_name`

Name	Description	Type	Dimensions
	are left in the working set of a process even if they are not in use. When free memory falls below a threshold, pages are trimmed from working sets. If pages are needed, they will be soft-faulted back into the working set before they leave main memory.		
process_working_set_bytes	Maximum number of bytes in the working set of this process at any point in time. If free memory is above a threshold, pages are left in the working set of a process even if they aren't in use. When free memory falls below a threshold, pages are trimmed from working sets. If pages are needed, they will be soft-faulted back into the working set before they leave main memory.	gauge	process_set, process_ppid, process_pid, process_cmdline, process_exe_path, process_service_name

17. Using Integrations

17.1. Cribl Edge To Cribl Stream

Cribl Edge automatically discovers logs, metrics, application data, etc. – in real time – from your configured endpoints, and delivers them to Cribl Stream or any supported destination. Meanwhile, Cribl Stream can help collect, reduce, enrich, transform, and route data from Cribl Edge to any destination. And using a Cribl TCP Source, you can collect and route data from Edge Nodes to Stream Worker Nodes connected to the same Leader, without incurring additional cost.

This guide outlines how to route data from an Edge Node (or an entire Fleet) to an existing Stream Worker Group for additional processing. We will walk you through the following:

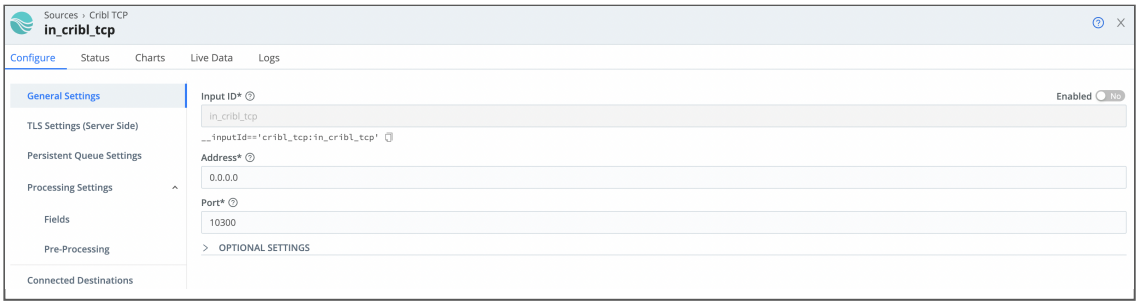
- Configure Cribl TCP Source on Cribl Stream to receive data from the Edge Node.
- Configure the Exec Source on Cribl Edge to collect data on the Edge Node.
- Configure the Cribl TCP Destination on Cribl Edge to send data to Cribl Stream.
- Configure a Route to Send the Data.

And finally, we will confirm the data flow.

While this use case connects Edge Nodes to Workers through the Cribl TCP Source and Destination, you can also use the Cribl HTTP Source and Destination in certain circumstances – such as when a firewall or proxy blocks raw TCP egress.

Configure the Cribl TCP Source on Cribl Stream

In Cribl Stream, start by configuring and enabling a Cribl TCP Source. The key requirement here is to set the Port to listen on. By default, the Cribl TCP Destinations listen on Port `10300`. To simplify our scenario, we will set the Cribl TCP Source to listen on the same Port. (Optionally, you can also configure TLS, Event Breakers, metadata fields, and/or a pre-processing Pipeline.)



When done, **Commit** and **Deploy** your changes. Before moving on to the next step, confirm that your Source is healthy.

Status of the Cribl TCP Source

On Cribl-managed Cribl.Cloud Worker/Edge Nodes, make sure that TLS is either disabled on both the Cribl TCP Source and the Cribl TCP Destination it's receiving data from, or enabled on both. Otherwise, no data will flow. In the Source, TLS is enabled by default.

Configure the Exec Source on Cribl Edge

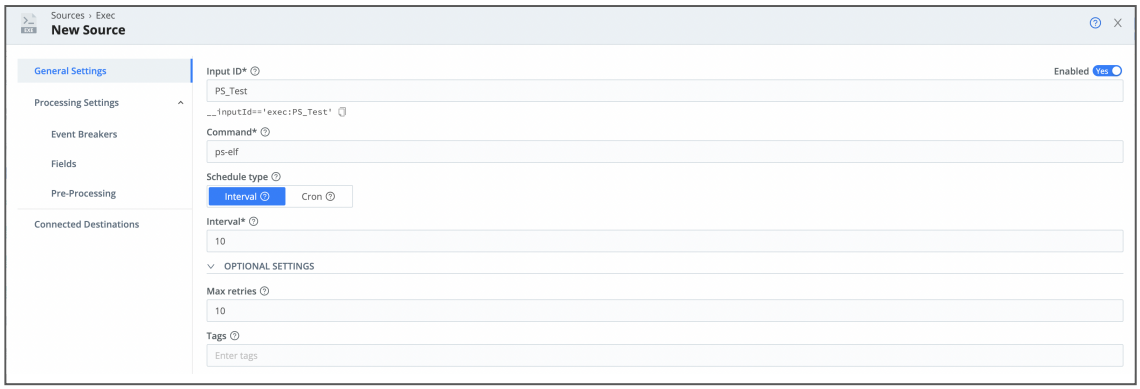
Next, we'll configure the Exec Source on your Edge Node. This Source will break the incoming streams of data into discrete events, and send them to Cribl Stream.

In this step, you can swap out the Exec Source by instead configuring a System Metrics or File Monitor Source. Or, configure multiple Sources to connect to the same Destination.

The Exec Source enables you to periodically execute a command and collect its `stdout` output. In the Exec Source's configuration modal, specify:

- Which command to execute.
- The number of times to attempt running the command.
- The interval between attempts.

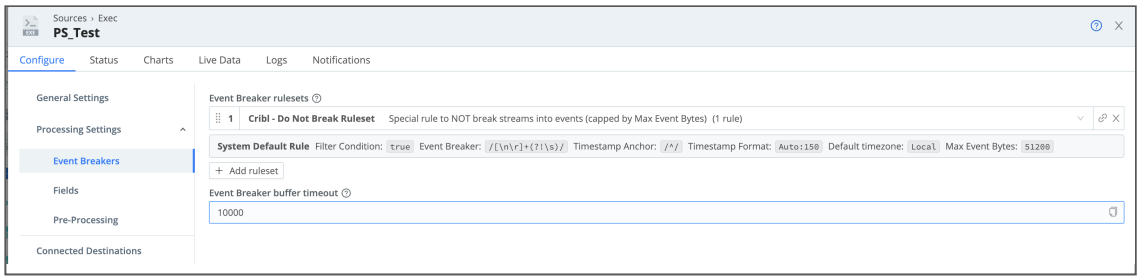
In our example, we are running the `ps` command to list and retrieve running processes every `10` seconds.



Configuring an Exec Source

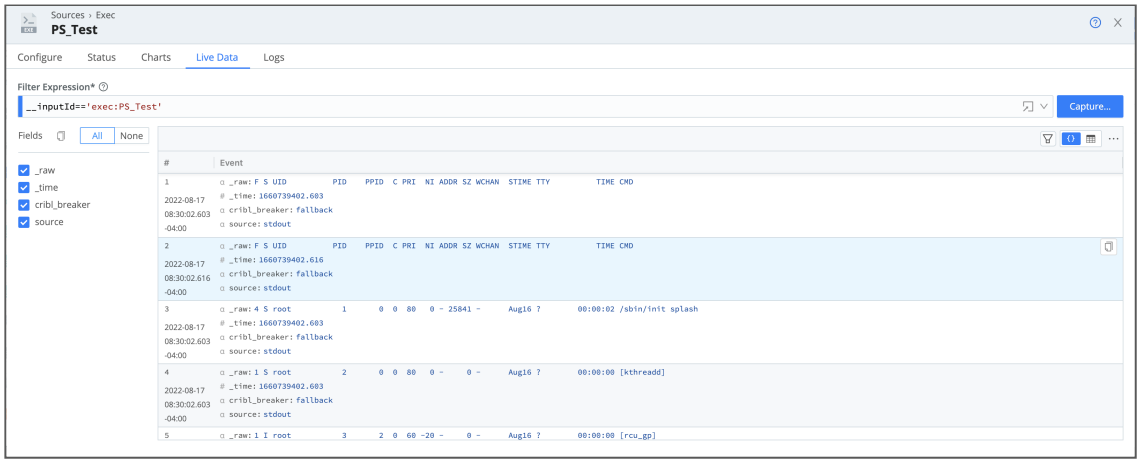
If we don't configure an Event Breaker, then with each capture we run on the dataset, each process will be ingested as its own event, without the header information. So to structure the data, we'll add an Event Breaker.

On the Exec Source configuration modal's left tab, select Event Breaker. In the **Event Breaker rulesets** drop-down, select `Cribl — Do Not Break Ruleset`.



Apply an Event Breaker

Next, preview your data on the modal's **Live Data** tab.



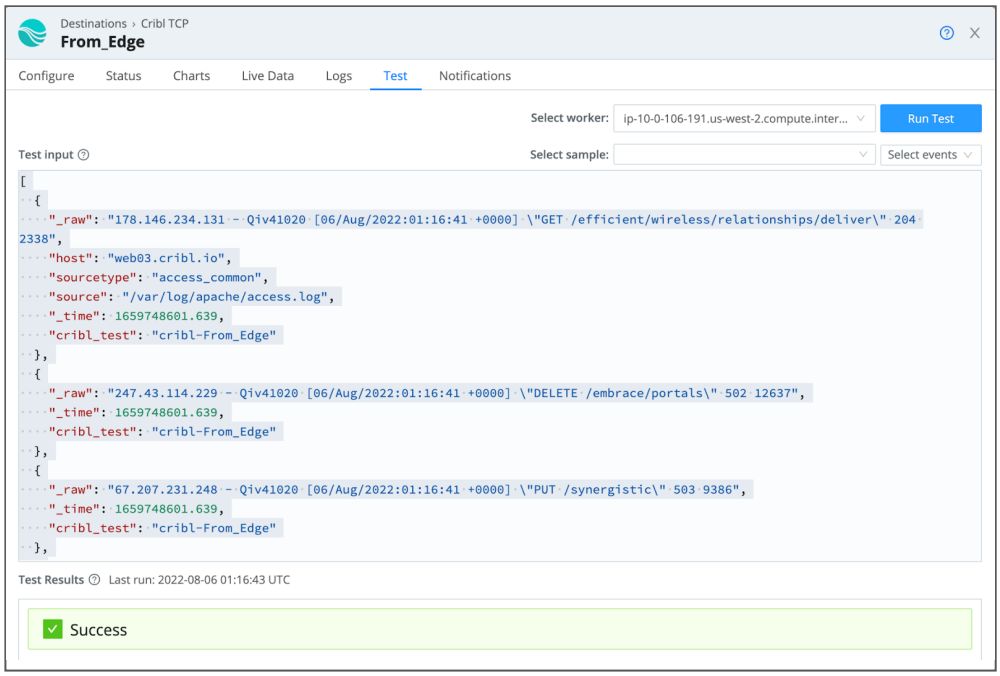
Preview Live Data

Configure the Cribl TCP Destination on Cribl Edge

To get the data flowing, we'll configure the Cribl TCP Destination on your Edge Node. A few things to note when configuring this Destination:

- Set the Port to listen on. For this example, we'll use the default `10300`. If you configure a different Port, make sure the Source points to the same Address and Port.
- If you don't have a load balancer in front of your Workers, you can configure load balancing directly on this Destination.
- Optionally, define your **Compression**, **Throttling**, and **Backpressure behavior** requirements.
- On Cribl-managed Cribl.Cloud Worker/Edge Nodes, make sure that TLS is either disabled on both the Cribl TCP Destination and the Cribl TCP Source it's sending data to, or enabled on both. Otherwise, no data will flow. In the Destination, TLS is disabled by default.

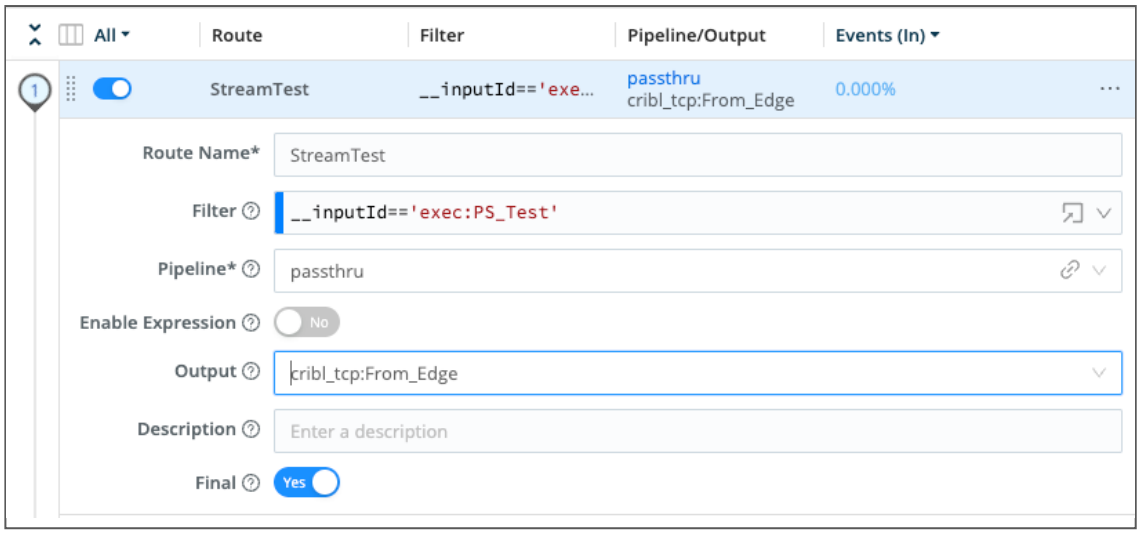
Once you've configured your Destination, test it to verify that your Edge Node can communicate with the Stream Worker Group.



Testing your Destination

Configure a Route to Send the Data

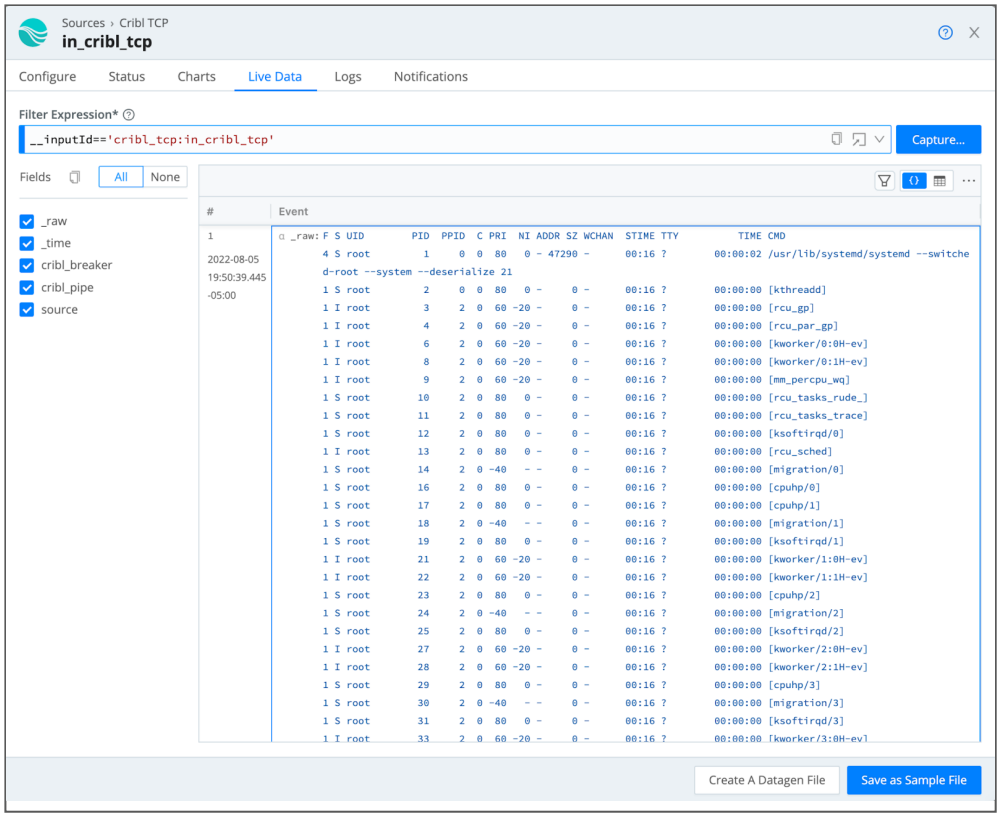
Finally, configure a Route to send your data to Cribl Stream. In this example, we are using the `passthru` Pipeline.



Routing your data

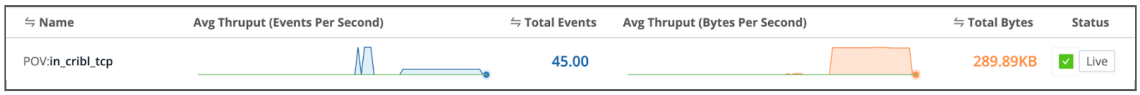
Confirm the Data Flow

To confirm that your data is flowing, navigate back to Cribl Stream's Cribl TCP Source. Run a **Live Data** capture on the Source.



Data flow in the Source

You can also check the Monitoring page's **Data** submenu, to isolate the throughput on your Source.



Monitoring the Source throughput

17.2. About Load Balancing

Cribl Edge will attempt to load-balance outbound data as fairly as possibly across all endpoints. For example, if FQDNs/hostnames are used as the Destination addresses, and each resolves to 5 (unique) IPs, then each Worker Process will have its # of outbound connections = {# of IPs x # of FQDNs} for purposes of the Destination.

Data is sent by all Worker Processes to a randomly selected endpoint, and the amount sent to each connection depends on these parameters:

1. Respective destination **weight**: If a **Load Weight** is set to `0`, Cribl Edge will not try to connect to that IP (nor to any IPs resolved for that FQDN). When a connection is blocked, Cribl Edge will apply a 10% penalty to the load weight, and will select the connection with the lower adjusted load.
2. Respective destination **historical data**: By default, historical data is tracked for 300s. Cribl Edge uses this data to influence the traffic sent to each destination, to ensure that differences decay over time, and that total ratios converge towards configured weights.

On Destinations that support load balancing, Cribl recommends enabling the feature, even if you only have one hostname – because this hostname can expand to multiple IPs. If you must disable load balancing, then to ensure robust connections, enable the alternative **Advanced Settings** > **Round-robin DNS** option. This will allow Cribl Edge to follow the DNS host's IP rotation, preventing connection pinning.

Example

Suppose we have two connections, A and B, each with weight of `1` (i.e., they are configured to receive equal amounts of data). Suppose further that the load-balance stats period is set at the default `300s` and – to make things easy – for each period, there are 200 events of equal size (Bytes) that need to be balanced.

Interval	Time Range	Events to be dispensed
1	*time=0s —> time=300s*	**200**

Both A and B start this interval with 0 historical stats each.

Let's assume that, due to various circumstances, 200 events are "balanced" as follows: `A = 120 events` and `B = 80 events` – a difference of **40 events** and a ratio of **1.5:1**.

Interval	Time Range	Events to be dispensed
2	*time=300s —> time=600s*	**200**

At the beginning of interval 2, the load-balancing algorithm will look back to the previous interval stats and carry **half** of the receiving stats forward. I.e., connection A will start the interval with **60** and connection B with **40**. To determine how many events A and B will receive during this next interval, Cribl Edge will use their weights and their stats as follows:

Total number of events: `events to be dispensed + stats carried forward = 200 + 60 + 40 = 300`. Number of events per each destination (weighed): `300/2 = 150` (they're equal, due to equal weight). Number of events to send to each destination `A: 150 - 60 = 90` and `B: 150 - 40 = 110`.

Totals at end of interval 2: `A=120+90=210`, `B=80+110=190`, a difference of **20 events** and a ratio of **1.1:1**.

Over the subsequent intervals, the difference becomes exponentially less pronounced, and eventually insignificant. Thus, the load gets balanced fairly.

If a request fails, Cribl Edge will resend the data to a different endpoint. Cribl Edge will block only if **all** endpoints are experiencing problems.

17.3. Destination Backpressure Triggers

This topic describes the conditions that can trigger backpressure in Cribl Edge Destinations. Backpressure exists when a destination receives more data than it can currently send. It typically results from network connectivity problems or when a Destination is receiving data faster than it can forward that data to the downstream service.

When a Destination signals backpressure, it does so to notify the Sources generating data that its buffers are full and that the incoming data flow should be paused until there is room for more data.

You can configure your desired behavior through a Destination's **Backpressure Behavior** drop-down. Where other options are not displayed, Cribl Edge's default behavior is **Block**. For details about all the above behaviors and options, see Persistent Queues.

HTTP-Based Destinations

HTTP-based Destinations trigger backpressure when they encounter errors while connecting or when they receive certain HTTP response codes. If the HTTP call returns a `500` series (`500`–`599`) HTTP code, the Destination adds the events in the buffer back to its internal buffer and tries to send them again. The retry attempt can also trigger backpressure, depending on the frequency at which this happens.

If the Destinaton receives a `400` series response code, it will drop (not retry) the events, because these response codes indicate an error in payload or formatting (for example). See this Mozilla page on the `400 Bad Request` code for more information.

Finally, HTTP-based Destinations trigger backpressure when their processing loads exceed the specified limit. HTTP-based Destinations handle requests by creating a connection, sending the request payload, and receiving and parsing a response. Because this process can take time, HTTP Destinations allow more than one request to be in progress at a time. When the number of requests in flight exceeds the value set in **Advanced Settings** > **Request concurrency**, the Destination will trigger backpressure.

The following conditions will trigger backpressure:

- Connection failure.
- Connection or request timeouts.
- Data overload – i.e., the Source sends more data than the Destination will accept.
- Reaching the limit on concurrent (in-flight) requests.
- HTTP `500` series responses.

You can set the number of in-flight or concurrent requests in **Advanced Settings** > **Request concurrency**.

HTTP-based Destinations include:

- Splunk HEC
- Elasticsearch
- Wavefront
- SignalFx
- Sumo Logic
- Honeycomb
- Datadog
- NewRelic Ingest Logs & Metrics
- NewRelic Ingest Events
- Azure Sentinel
- CloudWatch Logs
- Webhook
- Grafana Cloud
- Prometheus
- Loki
- Cribl HTTP
- InfluxDB
- Google Cloud Chronicle
- Google Cloud Logging

Filesystem-Based Destinations

Filesystem-based Destinations that send data to cloud-based services trigger backpressure when Cribl Edge can't move a file or upload it to the cloud.

Filesystem-based Destinations queue events in files on disk in a staging directory. When a batch of events is ready for transmission, Cribl Edge closes the file, optionally compresses it, and transmits the file to the downstream service.

For the Filesystem Destination, the batch simply moves from the staging directory to the output directory.

For other filesystem-based Destinations, Cribl Edge uploads the file to a cloud service, such as AWS S3, Azure Blob Storage, or other. For these Destinations, a failure to upload the file will trigger backpressure.

These scenarios include:

- Permissions problem writing to the staging directory – i.e., the user running the Cribl Edge process does not have permission to write to the staging directory.
- The filesystem for the staging or output directory is full (impacts Filesystem Destination only).
- The credentials provided for the cloud service are invalid.
- The Destination doesn't have the permissions required to upload to the cloud service.
- Cloud service permissions issue – i.e., upload fails because the cloud user (identified in credentials) does not have write permissions.
- Network connectivity – The Destination can't communicate with the cloud service due to network connectivity issues.

Filesystem-based Destinations include:

- Filesystem
- Azure Blob Storage
- Google Cloud Storage
- Amazon S3
- MinIO
- CrowdStrike Falcon LogScale
- Amazon Security Lake
- Amazon Data Lakes S3

TCP-Based Destinations

TCP Sources establish a persistent connection with a Destination and send data. These Sources maintain connection to the Destination and will re-establish that connection if they detect a close or write timeout.

If the connection fails, Cribl Edge uses a backoff algorithm to avoid spiking the CPU. This backoff logic makes the first retry at 2 seconds, then gradually increases the time between retries up to a maximum of 60 seconds.

The following conditions can trigger backpressure from a TCP-based Destination:

- Connection failure.
- Authentication failure.
- Write timeout, where a Source sends data but doesn't receive an `ack` at the TCP layer. In this case, Cribl Edge closes the connection to the Destination and tries to re-establish it.
- The Destination receives data from the Source faster than it can send data over the network to the downstream service, causing internal buffers to fill up. This condition can occur due to a slow network

connection or slow processing on the downstream service. When space in the internal buffers becomes available, backpressure is relieved, allowing the Source to send more data.

The following Destinations are TCP-based and adhere to the rules described above. Some Destinations use either TCP or UDP to send messages, but they trigger backpressure and use the backoff algorithm described above only when used with the TCP protocol:

- Splunk Single Instance
- Splunk Load-Balanced
- TCP JSON
- Cribl TCP
- Syslog (TCP or UDP)
- StatsD (TCP or UDP)
- StatsD Extended (TCP or UCP)
- Graphite (TCP or UCP)

Kafka

Kafka is a binary protocol that runs over TCP. The protocol defines three roles – producers, consumers, and brokers.

Producers send data to a Kafka topic (think of it as a message queue). Kafka consumers read data from Kafka topics. Brokers receive data from topic producers and send data to topic consumers – i.e., serve as go-betweens for producers and consumers.

Cribl Edge's Kafka-based Destinations are producers. They send data to Kafka brokers. Since the protocol is TCP-based, the conditions that trigger backpressure are similar to those of Cribl Edge's TCP Destinations.

The specifics for connection and request retries are defined by the Kafka JS library in conjunction with the configuration parameters defined for each Destination. In all cases, the Destinations have internal buffers that are sent to the Broker when they're considered full (on a size and time basis).

The following conditions can trigger backpressure for Kafka-based Sources:

- Failure to connect to the broker, lack of network connectivity, or the broker is down.
- Invalid credentials – Destination connects successfully, but the connection is rejected due to a failure to authenticate.
- Invalid permissions – The credentials assigned do not have proper permissions to send events to the topic.

- The Destination is attempting to send data faster than the broker can receive it. Backpressure will be relieved when internal buffers have room for more events.

Cribl Edge supports the following Kafka-based Destinations:

- Kafka
- Azure Event Hubs
- Confluent Cloud
- MSK

Kafka Retries and PQ

In some situations, Kafka Destinations may take a long time to engage the PQ (1 to 6 minutes by default).

This delay can result from the default settings used with the Kafka Destination. The relevant settings are all in the **Advanced Settings** UI:

- **Connection timeout (ms)**: Defaults to `10000`,
- **Request timeout (ms)**: Defaults to `60000`.
- **Max retries**: Defaults to `5`.

These default settings mean that Cribl Edge will retry five times after a failed connection, with a 10-second timeout between each attempt. It will also retry five files after a failed request, with a timeout of 60 seconds between each attempt.

For faster PQ engagement, update these values. You can use smaller values for **Connection timeout** and **Request timeout** to decrease the interval between retries. You can also use smaller values for **Max retries** or even set it to `0` to eliminate retries altogether.

UDP

UDP is a connectionless protocol, so there is no concept of backpressure at the network layer. As a result, UDP Destinations will never trigger backpressure.

The following Destinations support UDP. Except for SNMP Trap, these Destinations also support TCP, but the UDP versions do not trigger backpressure.

- SNMP Trap
- Syslog
- StatsD
- StatsD Extended

- Graphite

SQS

The SQS (Simple Queue Service) Destination uses the AWS SQS library to send events to the Destination's configured queue. Like Kafka, this Destination acts as a producer that publishes messages to the provisioned SQS queue.

This Destination has five buffers with 10 events per buffer for a total of 50 events that can be queued before backpressure is enabled. The reason that backpressure might be generated for this Destination include:

- Invalid credentials.
- Invalid permissions.
- Queue does not exist and **Create Queue** is disabled.
- Lack of network connectivity.

Amazon Kinesis

Kinesis runs over HTTPS and behaves like Pub/Sub Destinations.

This Destination serializes and publishes events in a newline-delimited JSON format. In the context of the Kinesis Destination, it behaves as a producer where a Kinesis source is considered a consumer. Since the protocol is HTTP-based, the conditions that trigger backpressure are similar to those of other Cribl Edge HTTP Destinations.

The Kinesis library defines the specifics for connection and request retries in conjunction with the configuration parameters defined for each Destination. In all cases, the Destinations have internal buffers that are sent to the Broker when they're considered full (on a size and time basis).

The following conditions can trigger backpressure for Kinesis-based Destinations:

- Failure to connect to the Kinesis service, network connectivity, or the service is down.
- Invalid credentials – Connects successfully but connection is rejected due to a failure to authenticate.
- Invalid permissions – The credentials assigned do not have proper permissions to send events to the Kinesis stream.
- The Destination is attempting to send data faster than the service can receive it. Backpressure will be relieved when internal buffers have room for more events.

Google Cloud Pub/Sub

Google Cloud Pub/Sub is similar to Kafka and Kinesis in that there are producers that send data and consumers that receive data. The Cribl Edge Google Cloud Pub/Sub Destination is a producer that sends data to the Google Cloud Pub/Sub service. Google Cloud Pub/Sub leverages gRCP, however, which uses protocol buffers as both its interface definition language (IDL) and as its underlying message interchange format.

The conditions that trigger backpressure for a Google Cloud Pub/Sub Destination are:

- Failure to connect to the Google Pub/Sub service, network connectivity, or the service is down.
- Invalid credentials – Connects successfully but connection is rejected due to a failure to authenticate.
- Invalid permissions – The credentials assigned do not have proper permissions to send events to the topic.
- The Destination is attempting to send data faster than the service can receive it. Backpressure will be relieved when internal buffers have room for more events.

OpenTelemetry

OpenTelemetry supports gRPC or HTTP to send messages. For the gRPC protocol, the same rules described for other TCP Destinations generally apply to OpenTelemetry. The key difference is that this Destination uses the OpenTelemetry library, which handles write timeouts and reconnection logic.

For the HTTP protocol, the same rules apply to our HTTP Destinations for retries based on connectivity or `500` series response codes.

Load-Balanced Destinations

HTTP-Based Load-Balanced Destinations

Some HTTP-based Destinations support two modes – single host and multiple load-balanced hosts.

With a load-balanced version of an HTTP Destination, Cribl Edge spreads events across all configured Destinations, using a weighted algorithm. Under normal circumstances, this algorithm spreads the load across the configured Destinations according to the weight assigned to the Destination.

When one of the load-balanced Destinations is out of service (due to a write timeout or `500` series response code - buffers full), Cribl Edge removes that Destination from the list of hosts and moves any in-flight buffers to the next available host. Cribl Edge will try to resend events to the downed host in the background, and it will resume sending data to that host once the connection is restored.

HTTP Destinations that support load balancing include:

- Splunk HEC
- Elasticsearch
- Cribl HTTP

TCP Load-Balanced Destinations

Some TCP-based Destinations support two modes – single host and multiple load-balanced hosts.

With a load-balanced TCP Destination, Cribl Edge spreads events across all configured Destinations using a weighted algorithm. Under normal circumstances, this algorithm spreads the load across the configured Destinations according to the weight assigned to each Destination.

When one of the load-balanced Destinations is out of service (due to a write timeout or connection failure), Cribl Edge removes that Destination from the list of hosts and moves any in-flight buffers to the next available host. Cribl Edge will try to reconnect to the downed host and will resume sending data to that host once the connection is restored.

TCP load-balanced Destinations can be configured with one or multiple receivers. If one or more receivers go down, Cribl Edge will continue sending data to any healthy receivers.

TCP load-balanced Destinations include:

- Splunk Load-Balanced
- TCP JSON
- Cribl TCP
- Syslog (TCP version only)

Output Router

The Output Router is a wrapper destination that groups together 1..*N* Destinations. It's used in conjunction with event filters to route data to a selected Destination. The Output Router Destination emits backpressure only when one of its downstream Destinations is blocking.

17.4. Persistent Queues

Cribl Edge's persistent queuing (PQ) feature helps minimize data loss if a downstream receiver or is unreachable or slow to respond, or (when configured on Sources) during backpressure from Cribl Edge internal processing. PQ provides durability by writing data to disk for the duration of the outage, and then forwarding it upon recovery.

Persistent queues can be enabled:

- On Push Sources.
- On Streaming Destinations. (Sources can also take advantage of a Destination's queue to keep data flowing.) For details, see Persistent Queues Support.

As a paid feature, persistent queues are available:

- On customer-managed (on-prem) deployments with at least a Standard license.
- On Cribl.Cloud with an Enterprise plan.

Persistent Queues Supplement In-Memory Queues

Persistent queues trigger differently on the Destination versus Source side.

Destination Side

On Cribl Edge Destinations, in-memory buffers help absorb temporary imbalances between inbound and outbound data rates. For example, if there is an inbound burst of data, a Worker Process will store events in Destination buffers, and will then output them at the rate that the downstream receiver can catch up.

Only when buffers are saturated will the Destination impose backpressure upstream. (This threshold varies per Destination type.) This is where enabling persistent queues helps safeguard your data.

Destinations Without PQ

You can configure each Destination's backpressure behavior to one of **Block**, or **Drop Events**, or (on Destinations that support it) **Persistent Queue**.

In **Block** mode, the output will refuse to accept new data until the receiver is ready. The system will back propagate block "signals" all the way back to the sender (assuming that the sender supports backpressure, too). In general, TCP-based senders support backpressure, but this is not a guarantee: Each upstream

application's developer is responsible for ensuring that the application stops sending data once Cribl Edge stops sending TCP acknowledgments back to it.

In **Drop** mode, the Destination will discard new events until the receiver is ready. In some environments, the in-memory queues and their block or drop behavior are acceptable.

PQ = Durability

Persistent queues serve environments where more durability is required (e.g., outages last longer than memory queues can sustain), or where upstream senders do not support backpressure (e.g., ephemeral/network senders).

Engaging persistent queues in these scenarios can help minimize data loss. Once the in-memory buffer is full, the Source or Destination will write its data to disk. Then, when the downstream receiver or Cribl process is ready, the Source/Destination will start draining its queue.

By default, after data flow is re-established, a Destination will forward events in FIFO (first in, first out) order - it will send out earlier queued events before newly arriving events. However, in Cribl Edge 4.1 and later, you can instead prioritize new events by disabling Destinations' **Strict ordering** control.

Source Side

Push Sources' config modals also provide a PQ option to supplement in-memory buffering. When you enable PQ, you can choose between two trigger conditions:

- `Always On` mode will use PQ as a disk buffer for **all** events.
- `Smart` mode will engage PQ only upon backpressure from downstream receivers or Cribl internal processes.

To learn more about PQ options on both the Source and Destination sides, see Planning for Persistent Queues. For implementation details, see Configuring Persistent Queues.

Persistent Queue Details and Constraints

Persistent queues are:

- Available on Push Sources.
- Available on the output side (i.e., after processing) of all streaming Destinations, with these exceptions: Syslog and Graphite (when you select UDP as the outbound protocol), Azure Data Explorer (when you select Batching mode), and SNMP Trap.

For specifics on both sides, see Persistent Queues Support.

- Implemented at the Worker Process level, with independent sizing configuration and dynamic engagement per Worker Process.
- With load-balanced Destinations (Splunk Load Balanced, Splunk HEC, Elasticsearch, TCP JSON, and Syslog with TCP), engaged only when **all of the Destination's receivers** are blocking data flow. (Here, a single live receiver will prevent PQ from engaging on the corresponding Destination.)
- On Destinations, engaged only when receivers are down, unreachable, blocking, or throwing a serious error (such as a connection reset). Destination-side PQ is not designed to engage when receivers' data consumption rate simply slows down.
- Drained when at least one receiver can accept data.
- Not infinite in size. I.e., if data cannot be delivered out, you might run out of disk space.
- Not able to fully protect in cases of application failure. E.g., in-memory data might get lost if a crash occurs.
- Not able to protect in cases of hardware failure. E.g., disk failure, corruption, or machine/host loss.
- TLS-encrypted only for data in flight, and only on Destinations where TLS is supported and enabled. To encrypt data at rest, including disk writes/reads, you must configure encryption on the underlying storage volume(s).

Beyond the failure scenarios above: If you turn off PQ on a Source or Destination (or shut down the whole Source/Destination) before the disk queue has drained, the queued events will be orphaned on disk.

17.4.1. Planning For Persistent Queues

The Insider's Guide

This page delves into:

- Use cases for implementing persistent queues (PQ).
- Details about PQ behavior on Sources and Destinations, with various configuration and tuning options.
- Choices you should make before enabling PQ.
- Implications of those choices for your system's requirements and performance.

For granular configuration procedures, see Configuring Persistent Queueing.

Source-Side PQ

Source-Side PQ can serve as a stopgap for upstream senders that don't provide their own buffering, or whose small buffers cannot queue for very long. This applies to several syslog senders and HTTP-based Sources.

Here, PQ prevents data loss when the Source is unable to handle backpressure for an extended period. Backpressure might be triggered by error conditions, or by degraded performance (slowdowns or short backups), in downstream Cribl resources or outside receivers.

Always On Versus Smart Mode

Where senders do not provide their own buffering, and where Destination-Side PQ is not enabled: It is a good practice to enable Push Sources' PQ option, in either `Smart` or `Always On` mode. But which mode should you choose?

`Always On` mode is a good match for highly valuable data (such as security data), and for data ingested via UDP. (Relevant UDP Sources are Raw UDP, SNMP Trap, Metrics, and Syslog with UDP enabled.)

`Smart` mode might be sufficient for lower-grade data, such as events that are overwhelmingly `debug`-level.

Always On Requirements

`Always On` mode provides the most reliable data delivery, but you should plan for its impacts on throughput, memory, and hosts' disk resources.

Throughput

Each Source configured in `Always On` mode imposes a significant performance penalty on Workers' data ingestion rate, as compared to Sources without PQ enabled. This primarily involves the overhead to write to the filesystem. As a starting point, assume that each event will be burdened with approximately 1 second of added latency.

Memory

`Always On` mode can also increase memory demands, as compared to Sources without PQ enabled. If you find memory to be a constraint, you can experiment with tuning the **Max buffer size** in each Source's UI, or the `maxBufferSize` configuration key. However, smaller buffers will reduce the efficiency of disk writes, adding more latency.

Disk

`Always On` also imposes more disk requirements than `Smart` mode. Cribl generally recommends that you allocate each Worker Node enough disk space for at least 1 day's worth of queued throughput. For example: If one Source is sending 1 TB/day to two Worker Processes (evenly weighted), you should provision at least 500 GB queue storage for each Worker Process. Expand this basic calculation for additional Sources, and expand or contract it based on the typical length of outages you experience.

Cribl also recommends that you deploy the highest-performance storage available. This prevents IOPS (input/output operations per second) and throughput bottlenecks for both read and write operations.

Compression

Enabling the PQ compression option conserves disk space, with a compression ratio of about 8:1. But you get these savings at the cost of worsening the performance hit. Navigate this trade-off based on which factor is most valuable in your deployment.

Smart Mode Details

`Smart` mode engages only when a Source receives an unhealthy signal from a downstream process (Pipeline, Destination down or blocked, etc.). When engaged, `Smart` mode's latency penalty is about the same as with `Always On` mode. But this latency is highly variable, because `Smart` mode engages intermittently. So expect less latency overall, with less predictability – spikes and valleys.

`Smart` mode's CPU performance penalty is about twice that of `Always On`, because of the overhead to switch queueing on and off. But again, this is highly variable, because performance will degrade only when

PQ engages/disengages. As with `Always On`, if memory is a constraint, you can experiment with reducing **Max buffer size** (UI) or `maxBufferSize` (config-file) values. Compression is supported in `Smart` mode, as in `Always On` modes.

Smart Mode Interaction with Multiple Destinations

If you enable `Smart` mode on a Source that's routed to multiple Destinations, you should plan around the unintended interaction that PQ can cause among these Destinations:

If at least one connected Destination sends an unhealthy signal, this Source will queue new events to disk even if some peer Destinations remain up. This will delay data delivery to the peer healthy Destinations. The same interaction can occur between Destinations grouped in a connected Output Router.

To prevent this unintended behavior, you can either:

- Disable PQ on the Source, and instead configure PQ on individual Destinations; or
- Configure separate Worker Groups to connect this Source to separate Destinations.

How Source PQ Works

Once enabled on a specific Source, in `Smart` mode, PQ will engage when any of the following occur:

- A connected Destination is fully down, and Destination-side PQ is not enabled.
- A connected Destination is slow, or blocking.
- A downstream Cribl Pipeline/Route is slow.

These triggers are not relevant if you've enabled Source PQ in `Always On` mode. In that case – it's always on.

When PQ engages on a Source:

- Events queue **before** any pre-processing Pipeline takes effect.
- Metrics (data in) will show `0` bytes/events while events are buffered to the disk queue.
- Cribl Edge Pipelines/Routes will process the data with a delay, when PQ later drains.

Because PQ always carries some latency penalty, it is redundant to enable PQ on a Source whose upstream sender is configured to safeguard events in its own disk buffer.

If you turn off Source PQ (or the whole Source) before its queues drain, any events still queued on disk will be orphaned.

Destination-Side PQ

You can enable PQ on supported Destinations to safeguard data during extended errors, and to smooth out CPU load spikes when a Destination recovers from a transient outage.

Unlike Source-side PQ, Destination-side PQ will engage only when the Destination is **completely** unavailable, not just slow.

Destination PQ is a good match for receivers like Splunk, where there are known outages. Because it has no `Always On` mode, Destination PQ imposes less overall performance overhead than Source PQ.

As with Source PQ, size your disk capacity according to your risk tolerance, but Cribl's initial recommendation is to size for at least 1 day's worth of storage on each Worker Node.

Fallback (Queue-Full) Behavior

Unlike Source PQ, on Destinations where you've enabled PQ, you also select a fallback **Queue-full behavior**.

With the default `Block` option, if a Destination's persistent queue fills up, the Destination will block back to the corresponding Source. (Remember that a Destination can be fed by multiple Sources.) If the Source has its own PQ option enabled, it will begin queueing data. If not, the block signal will go all the way back to connected senders.

If you instead set the **Queue-full behavior** to `Drop new data`, Cribl Edge will simply discard data intended for the unavailable Destination, but will still allow the data to proceed to parallel Destinations.

Good use cases for `Block`: A simple Cribl Edge Route relaying Universal Forwarder data to a Splunk receiver; or, any Destination where you absolutely **must not** lose data.

How Destination PQ Works

If a Destination does not have load balancing enabled, PQ will engage upon a blocking or unavailable signal from the downstream receiver.

But with a load-balanced Destination, PQ will engage only if **all** of the Destination's downstream receivers are down or blocked. If some individual targets in the LB Destination are up, Cribl Edge will attempt to discover those receivers and deliver data to them.

When PQ is enabled on a specific Destination:

- PQ will engage when the Destination is completely unavailable, but not when it is simply slow.

- Files are written to disk **after** any post-processing Pipeline takes effect. (This is the mirror image of Source PQ, which writes to disk **before** pre-processing Pipelines.)
- New incoming data still flows to **other** Destinations that are up.
- When the Destination recovers, the default draining behavior is FIFO (first in, first out) But you can toggle off this **Strict ordering** default– and doing so enables a configurable **Drain rate limit (EPS)**.
- A **Clear Persistent Queue** button is available on each Destination, in case you need to clear out the queue's contents. This is destructive – it recovers the disk space, but abandons the queued data. Treat it as a panic button.

Destination PQ Better Practices

- Cribl generally recommends enabling the PQ **Compression** option on Destinations. (Disable compression if you find that it adds excessive latency.)
- When draining large volumes after recovery from an outage, you might find that PQ is crushing its Destination. In this case, disable **Strict ordering**, and tune the **Drain rate limit (EPS)** to maintain reasonable throughput for new data.

Better Practices for Source and/or Destination PQ

If your Source(s) and Destination(s) both support PQ, which side should you enable?

Enabling Source-side PQ (in `Always On` mode) is the most versatile and reliable choice. Source PQ is further upstream than Destination PQ, and can safeguard data bound for multiple Destinations.

However, you might opt for Destination-side PQ if your senders provide their own buffering, or if you don't want to add the system resources that Source Always On mode requires.

Enabling **both** Source- and Destination-side PQ on a Route is likely overkill. It can also complicate diagnosing issues.

Warning for Auto-Scaling Environments

With **persistent** storage, events on disk survive a Cribl Stream/Edge or OS restart. But you cannot count on this with ephemeral/auto-scaling environments. Stream Workers must **not** be destroyed while data has not drained from a PQ.

PQ Status Notifications and Monitoring

With an Enterprise or Standard license, you can configure Notifications that will be sent when PQ files exceed a configurable percentage of allocated storage, or when a Destination reaches the queue-full state

described above.

These Notifications always appear in Cribl Edge's UI and internal logs. You can also send them to external systems. For setup details, see Destination-State Notifications.

Destination PQ Notifications

With Destination PQ enabled, you can configure Notifications around two trigger conditions. Cribl recommends that you create both:

- Destination Backpressure Activated. This will send Notifications whenever backpressure engages, causing blocked or dropped events. It will also send Notifications when a queue fills up, falling back to blocked or dropped events.
- Persistent Queue Usage (saturation). Note that you can set these Notifications for **multiple** saturation levels.

Source PQ Alerting

Currently, Cribl Edge does not support Notifications around Source PQ states. However, if you've enabled `Always On` mode, alerts that PQ has engaged would be pointless – because all of the Source's incoming data is automatically written to the disk queue.

Also, if desired: On Fleets or customer-managed (on-prem) Worker Groups, you can enable the internal CriblLogs Source, create a Route to forward Cribl's internal logs to a downstream service of your choice, and configure alerts from there.

Monitoring PQ

Cribl Edge's native Monitoring dashboards interact with PQ in particular ways. When PQ engages, assume that your sole sources of truth about Sources' or Destinations' throughput are the dashboards at:

- **Monitoring** > **System** > **Queues (Sources),** and
- **Monitoring** > **System** > **Queues (Destinations)**.

Other Monitoring resources become unreliable during queueing. These include the top-level **Monitoring** dashboard, and:

- **Monitoring** > **Data** > **Sources,** or
- **Monitoring** > **Data** > **Destinations**.

These will not accurately represent queued data because, once PQ engages, data will be queued to disk, rather than arriving into Pipelines and downstream resources (with Source PQ) or into the affected Destinations (with Destination PQ). While the data is queueing, the two above dashboards for:

- Affected Sources will show no events per second (EPS) or bytes ingested while their data is queueing.
- Affected Destinations will show EPS and bytes sent while their data is queueing, but these metrics can be misleading. Here, they measure data going into the queue, but not proceeding to downstream receivers.

Some other Monitoring quirks to keep in mind:

- Activity in the **Monitoring** > **System** > **Queues (Sources)** and **Monitoring** > **System** > **Queues (Destinations)** dashboards sometimes shows up only when queues drain.
- Cribl Edge currently does not provide a Monitoring dashboard for disk queues' own volumes.

These gaps provide another reason to enable PQ Notifications where available.

Remember that with **or** without PQ enabled, Monitoring dashboards do not show usage of integrations' in-memory buffers.

After the unhealthy condition resolves, and disk queues start to drain, dashboards will resume accurately representing delivery rates and volumes.

Metrics and Capture Limitations

Finally, while PQ provides many benefits, also note the following limitations:

- While events go to a Source-side queue, Cribl Edge does not record metrics about that Source's inbound data (events/second or bytes).
- While events go to a Destination-side queue, metrics reflect events/second and bytes going into that Destination's queue, not proceeding beyond Cribl Edge.
- Data captures might not work while a Destination is blocked.

17.4.2. Configuring Persistent Queues

The following sections outline configuring PQ for Cribl.Cloud Workers managed by Cribl, and then for on-prem and hybrid Cribl.Cloud Workers.

Configuring PQ for Cribl-Managed Cribl.Cloud Workers

Persistent queuing is supported for Cribl-managed Workers in Cribl.Cloud deployments with an Enterprise plan. Configuration is considerably simpler with Cribl-managed Workers than with hybrid and on-prem Workers (both covered below). You configure PQ individually on each Source and Destination that supports it.

Sources

1. Click a Source to open its configuration modal.
2. Select the **Persistent Queue Settings** tab.
3. Toggle **Enable Persistent Queue** to on.

On Cribl-managed Workers, this tab exposes only the **Enable Persistent Queue** toggle. When enabled, PQ is automatically configured in `Always On` mode, with a maximum queue size of 1 GB of data per PQ-enabled Source, per Worker Process. The 1 GB limit is on uncompressed inbound data, and no compression is applied to the queue.

These options are not configurable. If you want to finely configure the maximum queue size, compression, PQ mode, and other options, use a hybrid Group.

Destinations

1. Click a Destination to open its configuration modal.
2. Set **Backpressure behavior** to `Persistent Queueing`.

On Cribl-managed Workers, the resulting **Persistent Queue Settings** tab exposes only a **Clear Persistent Queue** button. Once enabled, PQ is automatically configured with a maximum queue size of 1 GB of uncompressed data per PQ-enabled Destination, per Worker Process. The 1 GB limit is on uncompressed outbound data, and no compression is applied to the queue. If the queue fills up, Cribl Edge will block data flow.

These options are not configurable. If you want to finely configure the maximum queue size, compression, queue-full fallback behavior, and other options, use a hybrid Group.

If you want to disable persistent queuing:

1. Wait for the queue to fully drain.
   (The **Clear Persistent Queue** button is available as a destructive fallback. This will free up disk space, but will do so by deleting the queued data **without** sending it on to receivers.)
2. In the config modal's **General Settings**, select a different **Backpressure behavior**.
3. Re-save the config.

Logs

Cribl internal logs for Cribl-managed Workers might include entries labeled `Revived PQ Worker Process`. These indicate where a Cribl-managed Worker was newly assigned to drain a persistent queue left over from a shut-down Worker Node. You can use these log events to troubleshoot Cloud PQ issues.

Configuring PQ for On-Prem and Hybrid Workers

Persistent queuing is supported for Workers in on-prem deployments with a Standard license, and for hybrid Cribl.Cloud Workers with an Enterprise plan. (Cribl-managed Cloud Workers have a simpler configuration.) You configure persistent queueing individually for each Source and Destination that supports it.

On Sources:

1. Click a Source to open its configuration modal.
2. Select **Persistent Queue Settings**.
3. Toggle **Enable Persistent Queue** to on.

On Destinations:

1. Click a Destination to open its configuration modal.
2. Set **Backpressure behavior** to `Persistent Queueing`.

These selections expose the additional controls outlined below.

For an example of optimizing several PQ throttling options listed below, see our How Does Persistent Queuing Work Inside Cribl Stream? blog post.

Source-Side PQ Only

Mode: Select a condition for engaging persistent queues.

- `Always On`: This default option will always write events to the persistent queue, before forwarding them to Cribl Edge's data processing engine.
- `Smart`: This option will engage PQ only when the Source detects backpressure from Cribl Edge's data processing engine.

Max buffer size: The maximum number of events to hold in memory before reporting backpressure to the sender and writing the queue to disk. Defaults to `1000`. (This buffer is per connection, not just per Worker Process – and this can dramatically expand memory usage.)

Commit frequency: The number of events to send downstream before committing that Stream has read them. Defaults to `42`.

Always On versus Smart Mode

In Cribl Edge 4.1 and later, Source-side PQ defaults to `Always On` mode when you configure a new Source. `Always On` offers the highest data durability, by minimizing backpressure and potential data loss on senders. However, this mode can slightly slow down data throughput, and can require you to provision more machines and/or faster disks.

If you are upgrading from a pre-4.1 version where `Smart` mode was the default, this change does not affect your existing Sources' configurations. However:

- If you create new Stream Sources programmatically, and you want to enforce the previous `Smart` mode, you'll need to update your existing code.
- You can optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Consider switching to `Always On` mode if you find `Smart` mode causing Source-side PQ to frequently engage, and you've configured PQ with a large **Max file size** (e.g., 1 GB). However, a second option here is to retain `Smart` mode, but reduce the **Max file size** to a value no higher than the default `1 MB`.

(Source-side PQ excessively triggers backpressure only when it reaches the last queued file **and** that file is large. Keeping this threshold low prevents that condition.)

Common PQ Settings (Source and Destination Sides)

Max file size: The maximum data volume to store in each queue file before closing it and (optionally) applying the configured **Compression**. Enter a numeral with units of KB, MB, etc. If not specified, Cribl Edge applies the default `1 MB`.

Max queue size: The maximum amount of disk space that the queue is allowed to consume, on each Worker Process. Once this limit is reached, this Source or Destination will stop queueing data; Sources will

block, and Destinations will apply your configured **Queue-full behavior**. Defaults to `5 GB`. Enter a numeral with units of KB, MB, GB, or TB. For details, see Optimizing Max Queue Size.

Queue file path: The location for the persistent queue files. Defaults to `$CRIBL_HOME/state/queues`. To this value, Cribl Edge will append `/<worker-id>/<output-id>`.

Compression: Codec to use to compress the persisted data, once a file is closed. Defaults to `None`; `Gzip` is also available.

Optimizing Max Queue Size

When you upgrade Fleets, and customer-managed (on-prem) and hybrid Groups, to Cribl Edge 4.4 and later: Any undefined PQ **Max queue size** values in pre-4.4 Source/Destination configs will be set to the new 5 GB default value. This is a guardrail to control queues' consumption of host machines' disk space.

Adjust this default to match your needs. If some of your integrations' persistent queues had an undefined **Max queue size**, queues larger than 5 GB will trigger PQ fallback behavior until you resize this limit.

Also in Cribl Edge 4.4 and later, this field's highest accepted value is defined in the new **Group Settings** > **General Settings** > **Limits** > **Storage** > **Max PQ size per Worker Process** field. The corresponding `limits.yml` key is `maxPQSize`. Consult Cribl Support before increasing that limit's 1 TB default.

The changes here do not affect any existing config with a defined **Max queue size**, and do not affect Cribl-managed Cloud Workers' nonconfigurable 1 GB **Max queue size**.

Max Queue Size with Compression

If you enable **Compression** and also enter a **Max queue size** limit, be aware of how these options interact:

- Cribl Edge currently applies the **Max queue size** limit against the **uncompressed** data volume entering the queue.
- With this combination, Cribl recommends that you set the **Max queue size** limit **higher** than the volume's total available disk space – again disregarding compression. This will maximize queue saturation and minimize data loss. (For details, see Known Issues.)

Destination-Side PQ Only

Queue-full behavior: Determines whether to block or drop events when the queue is exerting backpressure (because disk is low or at full capacity). **Block** is the same behavior as non-PQ blocking, corresponding to the

Block option on the **Backpressure behavior** drop-down. **Drop new data** throws away incoming data, while leaving the contents of the PQ unchanged.

Clear Persistent Queue: Click this "panic" button if you want to delete the files that are currently queued for delivery to this Destination. A confirmation modal will appear - because this will free up disk space by permanently deleting the queued data, without delivering it to downstream receivers. (Appears only after **Output ID** has been defined.)

This section's remaining controls are available in Cribl Edge 4.1 and later.

Strict ordering: The default `Yes` position enables FIFO (first in, first out) event forwarding. When receivers recover, Cribl Edge will send earlier queued events before forwarding newly arrived events. To instead prioritize new events before draining the queue, toggle this off. Doing so will expose this additional control:

- **Drain rate limit (EPS)**: Optionally, set a throttling rate (in events per second) on writing from the queue to receivers. (The default `0` value disables throttling.) Throttling the queue's drain rate can boost the throughput of new/active connections, by reserving more resources for them. You can further optimize Workers' startup connections and CPU load at **Fleet Settings** > Worker Processes.

Minimum Free Disk Space

For queuing to operate properly, you must provide sufficient disk space. In distributed mode, you configure the minimum disk space to support queues (and other features) on each Fleet, at **Fleet Settings** > **General Settings** > **Limits** > **Min free disk space**. If available disk space falls below this threshold, Cribl Edge will stop maintaining persistent queues, and data loss will begin. The default minimum is 5 GB.

17.4.3. Persistent Queues Support

This page identifies Cribl integrations that do, and don't, support persistent queues.

Persistent Queues Support by Source Type

The dividing line here is simple:

- Cribl Push Sources, which ingest streaming data, support PQ.
- Pull and Collector Sources, which ingest data intermittently and/or from static files, omit PQ support.

Persistent Queues Support by Destination Type

Persistent Queues support, behavior, and triggers vary by Destination type, as summarized below.

HTTP-Based Destinations

HTTP-based Destinations handle backpressure based on HTTP response codes. The following conditions will trigger PQ:

1. Connection failure.
2. HTTP `500` responses.
3. Data overload – sending more data than the Destination will accept.

HTTP `400` response errors will not engage PQ, and Cribl Edge will simply drop corresponding events. Cribl Edge cannot retry these requests, because they have been flagged as "bad," and would just fail again. If you see `400` errors, these often indicate a need to correct your Destination's configuration.

HTTP-based Destinations include:

- Amazon Cloudwatch Logs
- Amazon S3
- Amazon SQS
- Azure Blob Storage
- Azure Monitor Logs
- Cribl HTTP
- CrowdStrike Falcon LogScale

- Datadog
- Data Lake > S3
- Elasticsearch
- Google Chronicle
- Google Pub/Sub
- Grafana Cloud
- Honeycomb
- InfluxDB
- Loki (Logs)
- New Relic Ingest: Logs & Metrics
- New Relic Ingest: Events
- OpenTelemetry
- Prometheus
- SentinelOne DataSet
- SignalFx
- Splunk HEC
- Sumo Logic
- WaveFront
- Webhook

TCP Load-Balanced Destinations

TCP load-balanced Destinations can be configured with one or multiple receivers. If one or more receivers go down, Cribl Edge will continue sending data to any healthy receivers. The following conditions will trigger PQ:

1. Connection errors on **all** receivers.

As long as even one of the Destination's receivers is healthy, Cribl Edge will redirect data to that receiver, and will **not** engage PQ.

2. Data overload – sending more data than the Destination will accept.

TCP load-balanced Destinations include:

- Splunk Load Balanced
- TCP JSON

- Syslog

Destinations Without PQ Support

Filesystem-based Destinations, and Destinations that use the UDP protocol, do not support PQ. These include:

- Amazon S3 Compatible Stores
- Data Lakes > Amazon S3
- Data Lakes > Amazon Security Lake
- Azure Blob Storage
- Azure Data Explorer when in batching mode
- Filesystem/NFS
- Google Cloud Storage
- MinIO
- SNMP Trap
- StatsD (with UDP)
- StatsD Extended (with UDP)
- Syslog (with UDP)

Filesystem-based Destinations do not support PQ because they already persist events to disk, before sending them to their final destinations.

UDP-based Destinations do not support PQ because the protocol is not reliable. Cribl Edge gets no indication whether the receiver received an event.

Other Destinations

Other Destinations that write to a single receiver (without load balancing enabled) generally engage PQ based on these trigger conditions:

1. Connection errors.
2. Fail to send an event (for any reason).

17.5. Integrating With Other Services

17.5.1. Cribl Edge As A Sidecar Container In AWS ECS

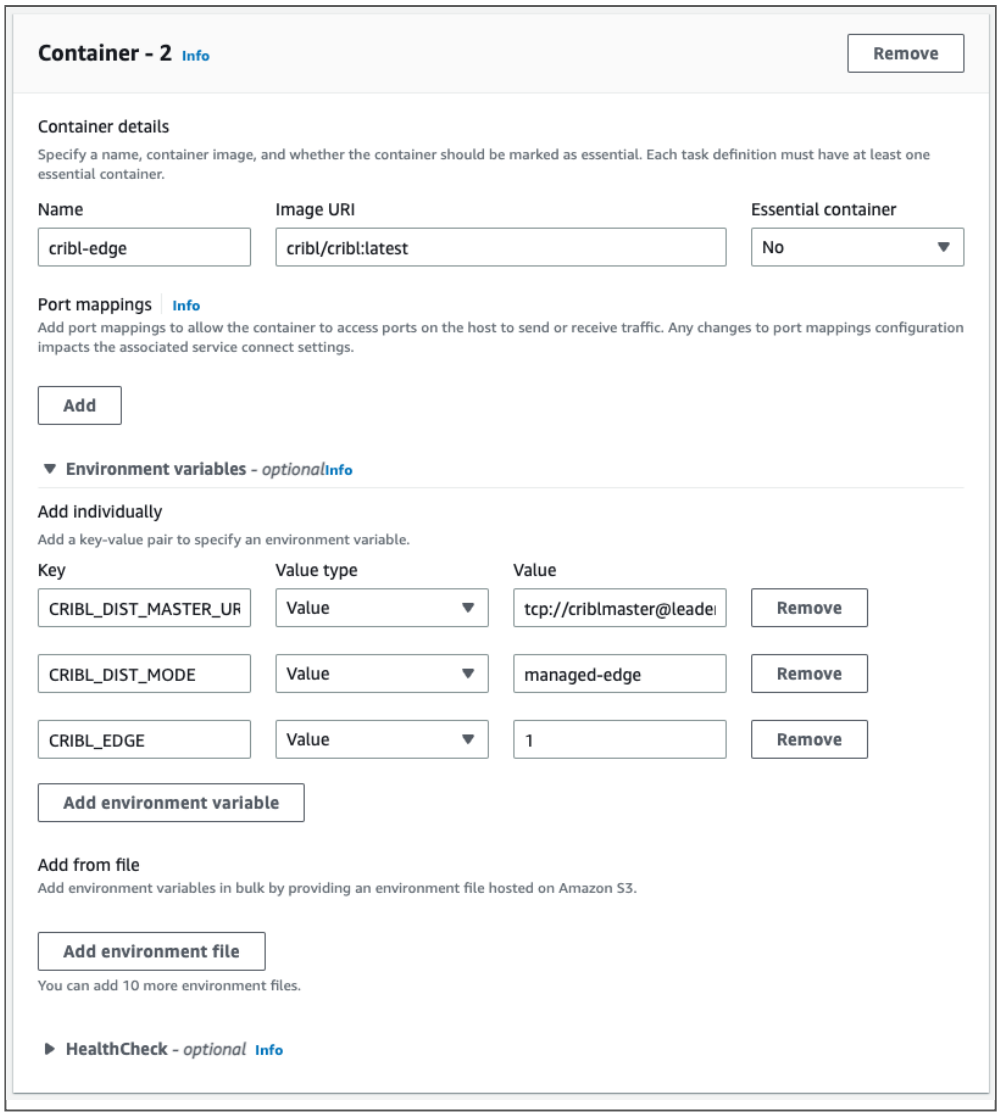
You can use Cribl Edge to collect logs from a service or task running in ECS. This is an alternative to forwarding logs to CloudWatch Logs.

Currently, you can't configure Windows hosts as sidecar containers in AWS.

Container Definition

Add a secondary container to the task definition. To configure the container, you must first set the following environment variables:

Name	Value
`CRIBL_DIST_MASTER_URL`	`tcp://<token>@<leader>:4200`
`CRIBL_DIST_MODE`	`managed-edge`
`CRIBL_EDGE`	`1`

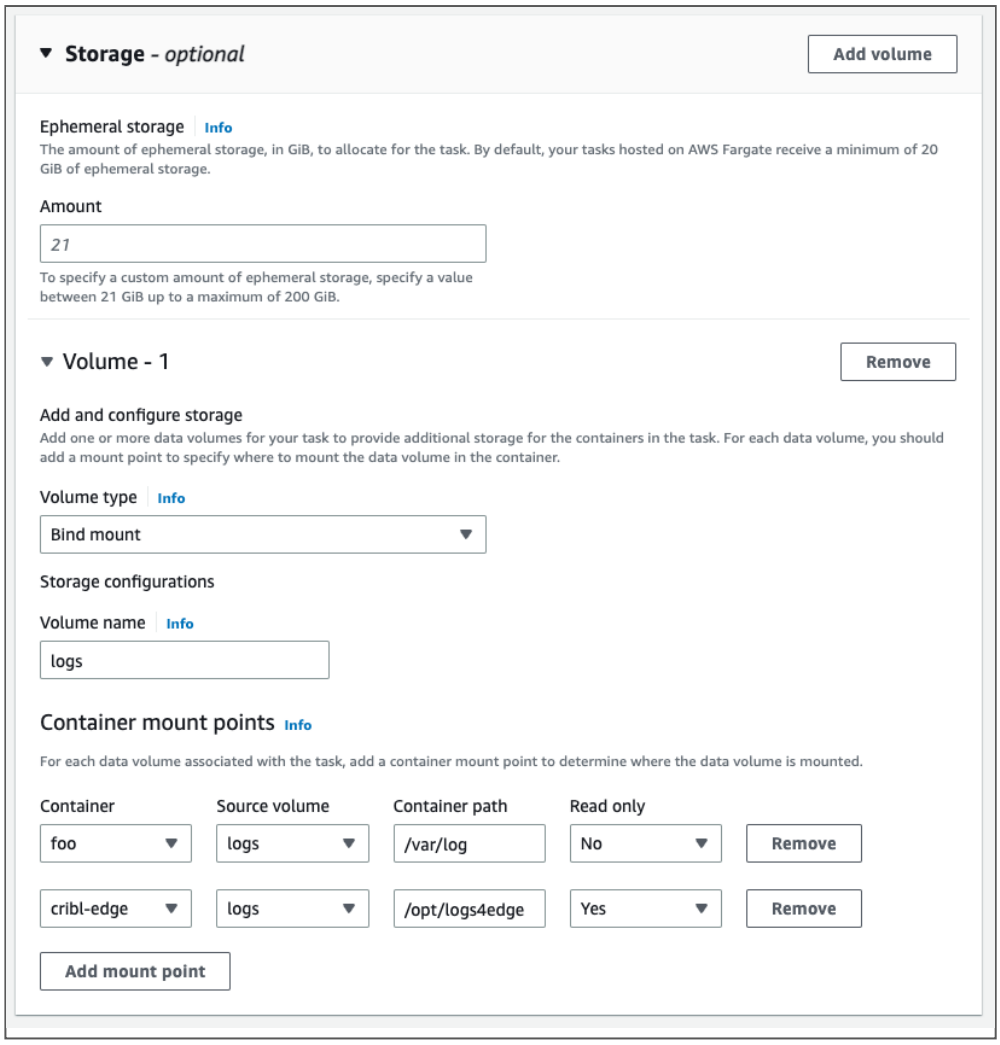


Add Environment Variables to Container Details

Configure Storage

To pass logs between containers, you can use a Bind Mount, which is a special folder mounted between containers. For details, see Using Data Volumes in Tasks.

Here's an example configuration from the ECS task definition:



ECS Task Definition Example

Sample JSON Container Definitions

You can adapt the following sample JSON definitions for the container. Note, this is not a full ECS task definition.

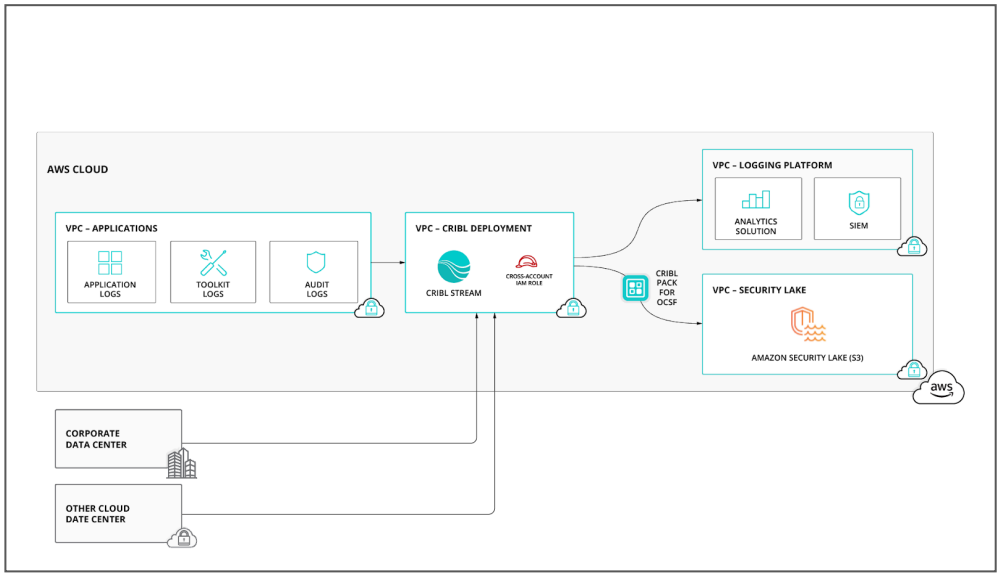
```
{
    "containerDefinitions": [
        {
            "name": "whatever",
            "image": "foo/bar:latest",
            ...
            "mountPoints": [
                {
                    "sourceVolume": "logs4edge",
                    "containerPath": "/var/log",
                    "readOnly": false
                }
            ],
            "volumesFrom": [],
            "logConfiguration": {
                "logDriver": "awslogs",
                "options": {
                    "awslogs-create-group": "true",
                    "awslogs-group": "/ecs/cribl-leader-plus-edge",
                    "awslogs-region": "us-west-2",
                    "awslogs-stream-prefix": "ecs"
                }
            }
        },
        {
            "name": "cribl-edge",
            "image": "cribl/cribl:latest",
            "cpu": 0,
            "portMappings": [],
            "essential": true,
            "environment": [
                {
                    "name": "CRIBL_EDGE",
                    "value": "1"
                },
                {
                    "name": "CRIBL_DIST_MODE",
                    "value": "managed-edge"
                },
                {
                    "name": "CRIBL_DIST_MASTER_URL",
                    "value": "tcp://cribledge@localhost:4200"
                `
```

```
                }
            ],
            "mountPoints": [
                {
                    "sourceVolume": "logs4edge",
                    "containerPath": "/opt/logs4edge",
                    "readOnly": true
                }
            ],
            "volumesFrom": [],
            "logConfiguration": {
                "logDriver": "awslogs",
                "options": {
                    "awslogs-create-group": "true",
                    "awslogs-group": "/ecs/cribl-leader-plus-edge",
                    "awslogs-region": "us-west-2",
                    "awslogs-stream-prefix": "ecs"
                }
            }
        }
    ],
    "volumes": [
        {
            "name": "logs4edge

            ",
            "host": {}
        }
    ],
   ... truncated ...
}
```

17.5.2. Amazon Security Lake Integration

Cribl Edge supports sending data to Amazon Security Lake as Parquet files that adhere to the Open Cybersecurity Schema Framework (OCSF). The diagram below illustrates the all-AWS variant of this architecture.



Integrating Cribl Edge with Amazon Security Lake

This guide walks you through an example where Cribl Edge, running in Cribl.Cloud, sends Palo Alto Networks logs to Amazon Security Lake. The workflow starts with a Cribl Edge Syslog Source, which ingests the Palo Alto data and routes it to a Cribl Edge Amazon Security Lake Destination, which then writes Parquet files to Amazon Security Lake. These Parquet files organize the data according to the OCSF schema for the appropriate OCSF Event Class, in this case Network Activity [4001].

You can work with other data sources besides Palo Alto Networks, and other OCSF Event Classes besides Network Activity [4001]. See Going Beyond This Example below.

If you want to use Amazon Security Lake with a Cribl Edge instance deployed on-prem or in your private cloud, the setup procedures will differ from those documented here. Please contact Cribl via the `#packs` channel of Cribl Community Slack.

Setting Up Amazon Security Lake

Complete the Amazon Security Lake Getting Started instructions, paying particular attention to the following actions:

1. Enable Amazon Security Lake for your AWS account.
2. Define the **collection objective** and **target objective**.

3. Select the regions where you want to create S3 buckets to store the data in OCSF format.

Note your Amazon Security Lake S3 bucket name for use later in the setup process.

Setting Up Cribl Edge

Complete the procedures in this section before doing anything else in Cribl Edge.

First, find and note the Amazon Resource Name (ARN) that enables AWS to locate your Cribl Edge instance.

1. Log in to Cribl.Cloud to open the Portal Page.
2. In the top nav, click **Network Settings** and open the **Trust** tab.
3. Click the copy icon next to the **Worker ARN**, and paste the ARN into a local text editor.
4. Copy the 12-digit number in the middle of the ARN. This is the AWS Account ID you'll need later in the setup process.

For example:

- If your ARN is `arn:aws:iam::222233334444:role/main-default`, then …
- Your AWS Account ID will be `222233334444`.

Next, install the Cribl Pack that Cribl Edge will use to send data to Amazon Security Lake in the required OCSF format. (As the reference shows, the Palo Alto Networks data chosen for this example requires the **OCSF Post Processor Pack**.)

1. In the top nav, click **Cribl.Cloud** to return to the Portal Page.
2. Click **Manage Stream** or **Manage Edge** to open the **Fleets** page.
3. Click the name of the Fleet whose Edge Nodes you want to send data to Amazon Security Lake. (For this example, we'll use the `default` Fleet.) This takes you to the **Manage** > **Overview** page.
4. Navigate to **Processing** > **Packs** and click **Add Pack**. From the drop-down, select **Import from Git** to open the **Import** modal.
5. For the **URL**, enter:

   ```
   https://github.com/asc-me-cribl/cribl_ocsf_postprocessing
   ```

6. For the **New Pack ID**, enter:

   ```
   Cribl-OCSF_Post_Processor
   ```

7. Click **Import** to close the modal. Cribl Edge will take a little time to finish importing the Pack.

At the bottom of the list of Packs, you should now see your newly imported Pack, listed with the display name of **OCSF Post Processor**.

Finally, you will need the Parquet schema that supports the OCSF Event Class you're working with. In this example, that's OCSF Event Class 4001, so as shown in the reference below, you'll need the `OCSF 4001 Network Activity` Parquet schema.

- First, download the Parquet schema from this GitHub repo.
- Then, upload the schema to the Cribl Edge Parquet schema library as described here.

Once this is done, the Parquet schema you need should be available when you configure the **Parquet Settings** for your Cribl Edge Amazon Security Lake Destination.

Please post any questions you have about these procedures to the `#packs` channel of Cribl Community Slack.

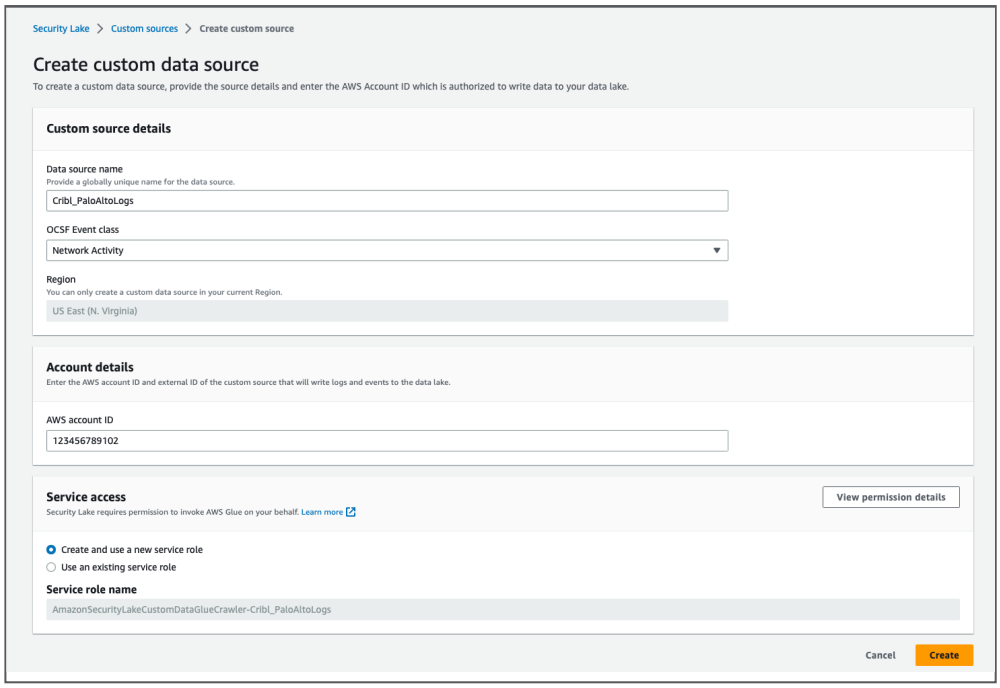
Creating an Amazon Security Lake Custom Source

Because Cribl is not a native Amazon service, you'll configure what Amazon calls a **Security Lake custom source**, and the Cribl Edge Amazon Security Lake Destination will write to that.

Before you begin:

1. Have your Cribl.Cloud AWS account ID, which you identified earlier, handy.
2. Ensure that you have permissions in AWS Lake Formation to create AWS Glue databases and tables. For more information, see Granting and revoking permissions using the named resource method in the AWS docs.

Then complete the instructions here in the AWS docs. These docs will take you through the **Create custom data source** template show below.



Creating an Amazon Security Lake custom source

Once the custom source has been created, still in AWS, navigate to **Security Lake** > **Custom sources** and you should see your new custom source in the list.

Now locate the **Provider role ARN**. This is the ARN that Cribl Edge will use to declare its role to AWS.

1. Click the copy icon next to the **Provider role ARN** to add it to your clipboard. Keep the ARN handy for use later in the setup process.
2. Navigate to the **Identity and Access Management (IAM)** screen.
3. From the left menu, select **Access management** > **Roles**.
4. In the **Roles** search box, paste the part of the ARN that begins `AmazonSecurityLake` and ends with the region.
5. When the role name appears below, click the link.
6. In the **Trust relationships** tab, view the `ExternalId` element in the JSON object that appears there.

This is the External ID for the trust relationship, which you created when went through the template. You will need this ID later in the setup process.

Creating a Cribl Amazon Security Lake Destination

Back in Cribl Edge, create a new Amazon Security Lake Destination in the Routing UI as described in the Destination topic. In the **New Destination** modal, configure the settings specified below. For all settings **not** mentioned in the following notes, you can keep the defaults.

General Settings

S3 Bucket name: The S3 bucket name you noted when setting Up Amazon Security Lake.

Region: Select the region where the S3 bucket is located.

Optional Settings

File name prefix expression: Keep the default (`CriblOut`) if it satisfies your requirements; otherwise, edit to suit your needs.

Authentication

Authentication method must be set to `Auto` (the default).

Assume Role

Cribl strongly recommends using the **Auto** authentication method in conjunction with the **Assume Role** settings as described below. Both Cribl and Amazon discourage the use of other approaches.

When using Assume Role to access resources in a different region than Cribl Stream, you can target the AWS Security Token Service (STS) endpoint specific to that region by using the `CRIBL_AWS_STS_REGION` environment variable on your Edge Node. Setting an invalid region results in a fallback to the global STS endpoint.

AssumeRole ARN: This is the **Provider role ARN** you copied after creating your custom source in Amazon Security Lake.

External ID: Enter the ID that you specified when creating your Amazon Security Lake custom source.

Processing Settings

Post-Processing

Pipeline: From the drop-down, select the Pack you installed earlier:

```
PACK Cribl-OCSF_Post_Processor (OCSF Post Processor)
```

System fields: Remove any fields specified here.

Parquet Settings

Parquet schema: From the drop-down select `OCSF 4001 Network Activity`, since OCSF Event Class 4001 describes the events coming from Palo Alto Networks in this example.

At this point, you can click **Save**; the defaults for **Advanced Settings** should work fine for this example. Then, **Commit** and **Deploy**.

You must create one Cribl Edge Amazon Security Lake Destination for **each unique pairing** of a data source with an OCSF Event Class.

- In this example, we're ingesting Palo Alto Networks Threat and Traffic data, which requires the OCSF Network Activity [4001] Event Class. That's one pairing.
- If you also wanted to ingest CrowdStrike Network Activity data, that would be a new data source paired with the same OCSF Event Class – i.e., **that would be a second unique pairing**. You would need to create a **separate** Cribl Edge Amazon Security Lake Destination for that pairing.

The reference below shows all the possible pairings of data sources and OCSF Event Classes.

Connecting a Cribl Source to the S3 Destination

We'll now configure a Cribl Edge Syslog Source to ingest Palo Alto Networks system logs, using the QuickConnect UI. (The reference below shows what Cribl Edge Source to use for each supported data source.)

Navigate to QuickConnect as described here. After clicking the Syslog Source tile, click **Select Existing**, then click the pre-configured `in_syslog` Source. When prompted to `Switch in_syslog to send to QuickConnect instead of Routes?`, click `Yes`.

Click + and drag the connection line to the S3 Destination you created above.

The connection between the Syslog Source and the S3 Destination should now be enabled.

Commit and **Deploy** the changes.

Sending Data to Amazon Security Lake

Return to your AWS Console. You should see Parquet files landing in the S3 bucket. These files should contain the Palo Alto Networks syslog data you sent through Cribl Edge.

Troubleshooting and Testing

Good things to double-check:

- Are you sending the events to your Amazon Security Lake in Parquet format?
- Are your permissions set properly for your IAM role to write to the S3 bucket?

If the answer to either question is "No," your data will not make it into Amazon Security Lake.

To send sample events from the GitHub repo, using netcat:

1. Download the `oYLbFU.json` sample file.
2. Run the following command, replacing `<cribl_org_id>` with your Cribl.Cloud Organization ID:

```
cat oYLbFU.json | nc default.main-<cribl_org_id>.cribl.cloud 10070
```

With a Cribl.Cloud Enterprise plan, generalize the above URL's `default.main` substring to `<group-name>.main` when sending to other Fleets.

Going Beyond This Example

If you want to send data from sources other than Palo Alto Networks, you can adapt the above instructions to use the appropriate Source and Pack (and/or modifications to the OCSF Post Processor Pack) in Cribl Edge. This holds true as long as the OCSF Event Classes you want to work with are among those supported by Cribl Edge. For each unique pairing of a data source with an OCSF Event Class, you'll need to create a separate Amazon Security Lake Destination.

In the next section is a list of the supported data sources, what OCSF Event Classes they handle, and what Packs and Parquet schemas you need to work with them.

Reference: Supported Data Sources

Data Source	OCSF Event Classes Handled	Cribl Source	Cribl Pack Display Name	Parquet Schema(s)

Azure Audit Logs	3002, 3004	REST Collector against Microsoft Graph API	Azure Audit Logs to OCSF	`OCSF 3002 Authentication`
				
				`OCSF 3004 Entity Management`

Azure NSG Flow Logs	4001	Azure Blob Storage Source, run as scheduled jobs, not as a Pull Source	Azure NSG Flow Logs	`OCSF 4001 Network Activity`
Cisco ASA	4001	Syslog	Cisco ASA	`OCSF 4001 Network Activity`
Cisco FTD	4001	Syslog	Cisco FTD	`OCSF 4001 Network Activity`
CrowdStrike Account Change	3001	CrowdStrike FDR	Crowdstrike FDR Pack	`OCSF 3001 Account Change`
CrowdStrike Authentication	3002	CrowdStrike FDR	Crowdstrike FDR Pack	`OCSF 3002 Authentication`
CrowdStrike Network Activity	4001	CrowdStrike FDR	Crowdstrike FDR Pack	`OCSF 4001 Network Activity`
GCP Audit Logs Account Activity	3001	Google Cloud Pub/Sub	Google Cloud Audit Logs	`OCSF 3001 Account Change`
Palo Alto Networks (PAN) Threat and Traffic	4001	Syslog	OCSF Post Processor Pack	`OCSF 4001 Network Activity`

SentinelOne	1001, 2001, 3002, 4001	Amazon S3	SentinelOne Cloud Funnel	`OCSF 1001 File System Activity`
				
				`OCSF 2001 Security Finding`

Data Source	OCSF Event Classes Handled	Cribl Source	Cribl Pack Display Name	Parquet Schema(s)

				`OCSF 3002 Authentication`
				
				`OCSF 4001 Network Activity`

Windows Logon Activity	3002	Splunk TCP	Splunk Forwarder Windows Classic Events to OCSF	`OCSF 3002 Authentication`
ZScaler Firewall and Weblogs	4001	Syslog	OCSF Post Processor Pack	`OCSF 4001 Network Activity`

17.5.3. Amazon Elastic Kubernetes Services (EKS) Add-On For Edge

Amazon Elastic Kubernetes Services Amazon EKS is a managed container service to run and scale Kubernetes applications in the AWS Cloud.

Cribl Edge provides an optimized bundle for Amazon EKS cluster cost visibility. This means you can accurately track costs by namespace, cluster, Pod, or by organizational concepts such as team or application. Kubernetes platform administrators and finance leaders can use Cribl Edge to visualize a breakdown of their Amazon EKS cluster charges. This can help them to allocate costs and chargeback organizational units such as application teams.

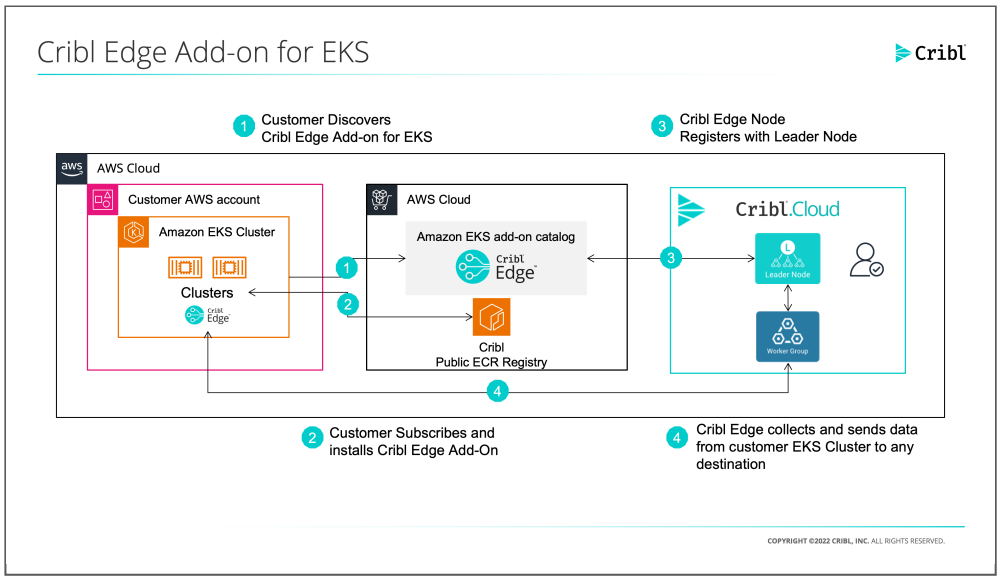
You can use your existing AWS support agreement to obtain support.

In This Article

In this article, you will learn more about how the Amazon EKS architecture interacts with Cribl Edge. You will also learn how to deploy Cribl Edge on EKS using one of two different methods:

1. Deploy Cribl Edge on an Amazon EKS cluster using the Amazon EKS add-on.
2. Deploy Cribl Edge on an Amazon EKS cluster via AWS CLI.

Architecture Overview



Amazon EKS architecture on Edge

Deploy Cribl Edge on an Amazon EKS Cluster

This section covers how to deploy Edge on an AWS EKS cluster.

Prerequisites

To deploy Edge on an Amazon EKS cluster using the Amazon EKS add-on, make sure you have:

- A Cribl.Cloud deployment or a self-hosted Cribl Leader.
- The ability to subscribe to Cribl Edge on AWS Marketplace.
- `kubectl`, AWS CLI, and (optional) `eksctl` installed on your machine.
- Access to an Amazon EKS Cluster.

Set up the Cribl Edge API Server

1. Log into your Cribl Deployment in a self-hosted deployment, or log into Cribl.Cloud, and click **Manage Edge**.
2. In Edge, click the Fleet for which you want to access deployment details.
3. To update the API Server Settings, click **Manage** then **Fleet Settings** in the submenu.

Update the Host IP

On your Leader Node, update the Host IP in the API Server Settings:

1. In **Fleet Settings**, click **System** > **General Settings**.
2. Under **API Server Settings**, click **General**.
3. In the **Host** field, replace the default value with `0.0.0.0`.
4. Click **Save**.
5. Click **Commit & Deploy**.

Access the Cribl URI and Token

To enable your Cribl Edge add-on for EKS, you'll need the Cribl URI and token. You can collect it from the Kubernetes script in Edge:

1. In Edge, navigate to **Manage** > **Overview**.
2. Click **Add/Update Edge Node** and select **Kubernetes**.

3. In the **Script** field, copy the `cribl.leader` URI, starting at `tls:`. This contains the auth token and the organization name for your deployment.
   Example URI:
   `tls://<auth_token>@<cribl_org_id>.cribl.cloud?group=<default_fleet>`
4. Save this URI somewhere you can easily access it in next steps.

Enable the Cribl Edge Add-On from EKS Console

To get started, access your Amazon EKS console and open your EKS clusters.

1. In the **Add-ons** tab, select **Get more add-ons**.
2. Type `Cribl Edge` into the search bar.
3. Follow the onscreen instructions to enable the Cribl Edge add-on for your Amazon EKS cluster. Learn more about direct deployment to Amazon EKS clusters from this AWS blog post.

Configure AWS EKS Settings

After you enable the Cribl Edge add-on, configure its settings:

1. Under **Version**, select the latest version.
2. Under **Select IAM role**, select **Inherit from node**.
3. Twirl open **Optional configuration settings** and navigate to **Configuration values**.
4. Insert the following JSON configuration:

```
{
 "cribl": {
     "leader": "tls://<token>@<cribl_cloud_leader>.cribl.cloud?group=<default_
 }
}
```

5. Replace the `leader` value with the URI for your Edge Leader gathered in Access the Cribl URI and Token.
   - For self-hosted Cribl deployments, enter the publicly available URI and token for your Leader Node.
6. In **Conflict resolution method**, select **None**.

Enable Cribl Edge Add-on Using AWS CLI

You can also enable the Cribl Edge add-on via CLI.

Enter this on the command line:

```
aws eks create-addon --addon-name cribl_cribledge --cluster-name $YOUR_CLUSTER_NAME
--region $AWS_REGION --configuration-values file://path_to_cribl_leader.json
```

Example output:


```
{
    "addon": {
        "addonName": "cribl_cribledge",
        "clusterName": "$YOUR_CLUSTER_NAME",
        "status": "CREATING",
        "addonVersion": " v4.3.1-eksbuild.1",
        "health": {
            "issues": []
        },
        "addonArn": "arn:aws:eks:$AWS_REGION:xxxxxxxxxxxx:addon/$YOUR_CLUSTER_NAME/
        "createdAt": "2022-12-01T12:18:26.497000-08:00",
        "modifiedAt": "2022-12-01T12:50:52.222000-08:00",
        "tags": {}
    }
}
```


To monitor the installation status, you can run the following command:

```
aws eks describe-addon --addon-name cribl_cribledge --cluster-name
$YOUR_CLUSTER_NAME --region $AWS_REGION
```

Example output:

```
{
    "addon": {
        "addonName": "cribl_cribledge",
        "clusterName": "cribl",
        "status": "ACTIVE",
        "addonVersion": "v4.3.1-eksbuild.1",
        "health": {
            "issues": []
        },
        "addonArn": "arn:aws:eks:$AWS_REGION:xxxxxxxxxxxx:addon/cribl/cribl_cribled
        "createdAt": "2023-11-09T08:28:35.212000-05:00",
        "modifiedAt": "2023-11-09T08:29:44.401000-05:00",
        "tags": {},
        "configurationValues": "{\r\n\t\"cribl\": {\r\n\t     \"leader\": \"tls://
    }
}
```


Disable the Cribl Edge add-on

To disable the Cribl Edge add-on, you can run the following command:

```
aws eks delete-addon --addon-name cribl_cribledge --cluster-name $YOUR_CLUSTER_NAME
–region $AWS_REGION
```

More Information

If you need more help, see Amazon EKS Add-Ons documentation.

18. Troubleshooting

18.1. Known Issues

This page lists known issues affecting Cribl Stream and/or Cribl Edge.

2024-03-20 v.4.5.1 – Permissions error appears when you try to edit a Pipeline [CRIBL-23627]

Problem: Cribl Stream Project Members with the Editor permission cannot view or edit Pipelines within the Stream Project.

Workaround: None.

Fix: In Cribl Edge 4.6.

2024-03-18 – v4.5.1 – An error can occur when you click Save to create a new Worker Group or Fleet [CRIBL-23496]

Problem: In Edge and Stream, an error can occur when you click **Save** to create a new Worker Group or Fleet. This error appears as a warning: `The Config Helper service is not available because a configuration file doesn't exist or the settings are invalid. Please fix it and restart Cribl server.`

Workaround: Refresh the page.

Fix: In Cribl Edge 4.6.

2024-03-14 – v4.5.1 – Amazon S3 Source and Destination, Amazon SQS Destination, and CrowdStrike Source crash in FIPS mode [CRIBL-23436]

Problem: When Cribl Edge in FIPS mode attempts to read or write files, the Amazon S3 Source and Destination, Amazon SQS Destination, and CrowdStrike Source crash with a `digital envelope routines::unsupported` error. This happens because the underlying Amazon AWS SDK v2 is configured to use the MD5 algorithm, which is not allowed in FIPS mode. Cribl Stream 4.6.0 fixes this problem by configuring the AWS SDK to skip the checksum computation that used MD5.

Workaround: None.

Fix: In Cribl Edge 4.6.

2024-03-13 – v4.4.0-v.4.5.1 – Diagnostic bundles cannot be created when git is enabled [CRIBL-23374]

Problem: Attempting to create a diagnostic bundle fails when git is enabled, resulting in an error.

Workaround: In the **Create & Export Diag Bundle** window, disable **Include git**.

Fix: In Cribl Edge 4.6.

2024-03-12 v.4.4.4-v.4.5.0 – S3 Collector Path extractors setting has no effect [CRIBL-23338]

Problem: In the S3 Collector, the **Collector Settings** > **Path extractors** setting has no effect because of a defect in the `index.js` file.

Workaround: Contact Cribl Support for assistance replacing the `index.js` file with a corrected one.

Fix: In Cribl Stream 4.5.1.

2024-02-13 – v4.5.1 – Edge Leaders running 4.5.1 can't use the Legacy upgrade method to upgrade 4.5.0 Nodes when jobs/tasks are disabled [CRIBL-22801]

Problem: Edge Leaders on version 4.5.1 that have Legacy upgrades enabled and **Enable Jobs/Tasks** toggled to `No` cannot upgrade their 4.5.0 Edge Nodes. The job finishes with the task timing out, and the Edge Node is not upgraded.

Workaround: On the 4.5.1 Leader, enable Legacy upgrades **and** toggle **Enable Jobs/Tasks** to `Yes`. Commit and deploy as necessary. The Leader will be able to create a job with the upgrade task and upgrade its Edge Nodes to 4.5.1.

Fix: Version TBD.

2024-02-12 – v4.5.0 – The API Reference automatic authorization doesn't work in Cribl.Cloud [CRIBL-22822]

Problem: In Cribl.Cloud, the API Reference (under **Global Settings**) uses an invalid access token. Because of this, you will receive an HTTP 401 error when you attempt to run an API command.

Workaround: To obtain a valid token, do the following:

1. Open the Developer Tools network tab in your browser and copy the authorization bearer token from one of your requests. Alternatively, you can follow the docs steps here to create a Cribl.Cloud API credential and obtain a token. Copy the token.
2. On the API Reference page, click **Authorize**.
3. In the resulting modal, click **Logout**.
4. Enter your copied token in the **Value** field and click **Authorize**.

Fix: In Cribl Edge 4.5.1.

2024-02-12 – v4.5.0 – When in the context of a Pack, the Pipeline link in the routing table is broken [CRIBL-22762]

Problem: In the Pack context, under **Routes**, if you click the hyperlinked Pipeline name in the table, you will get an `Item not found` error. Clicking `Try again` from the error page does not resolve the issue. However, the link icon at the end of the Pipeline field in the routing table works as expected and brings you to the intended Pipeline page.

Workaround: Use the link icon at the end of the Pipeline field in the Route to access the Pipeline.

Fix: In Cribl Edge 4.5.1.

2024-02-09 – v4.5.0 – OTLP Metrics function does not parse `__criblMetrics` when metric values are strings [CRIBL-22731]

Problem: Documentation for the Publish Metrics Function indicates that it evaluates the **Metric Name Expression** to the **Event Field Name**, but it actually returns a `null`. As a result, the OTLP Metrics function is unable to properly parse `__criblMetrics` when the metrics values are strings.

Workaround: When using Publish Metrics and OTLP Metrics together, use the Numerify Function before the OTLP Metrics Function.

Fix: In Cribl Edge 4.5.1.

2024-02-09 – v4.5.0 – Fleet-level log searches no longer work [CRIBL-22716]

Problem: The **Edge** > **Manage** > **Logs** UI no longer supports searching logs across entire Fleets. The option will be removed in an upcoming release.

Workaround: Search individual nodes or across Fleets in Cribl Search instead.

Fix: In Cribl Edge 4.5.1.

2024-01-30 – v4.1.0 through Current – Product Documentation: Reference Architecture image download link is broken [CRIBL-22502]

Problem: The editable diagram that appears in the All-in-One Reference Architecture section of the documentation is broken for all non-current release versions. Clicking the link for the `SVG` or `draw.io` file download results in a 403 error.

Workaround: Use the links in the most recent version of the documentation.

Fix: Version TBD.

2024-01-19 – v4.4.0-v4.4.4 – Unable to upgrade Leader from 4.3.x to 4.4.x when an external KMS is enabled [CRIBL-22299]

Problem: When attempting to upgrade the Leader from 4.3.x to 4.4.x with an external KMS enabled, you will receive several `Secret failed to decrypt` errors and the web UI will fail to load. This is due to an issue with how `client.secret` is handled when an external KMS is enabled.

Workaround: None.

Fix: In Cribl Edge 4.5.

2024-01-19 – v3.1.0 through Current – Notification targets created in one Cribl product can be deleted from another without warning [CRIBL-22292]

Problem: When you create a Notification target, it will be made available to all products in the Cribl suite. When you attempt to delete a target from the same application you created it in (for example, Cribl Stream), the product will properly prevent the deletion when there is an active notification using the target. However, if you switch to a new application (for example, Cribl Search), you will be allowed to delete the target without error or warning. This will leave the active notification in the original product (Stream) with an undefined target. You should not be able to delete the target if it is used by any active notification in any Cribl suite product.

Workaround: None.

Fix: In Cribl Edge 4.5.1.

2024-01-16 – v4.4.4 – Already-existing Webhook Destinations show unexpected UI behavior upon upgrade to v4.4.4 [CRIBL-22184]

Problem: If you have a Webhook Destination in a Cribl Edge deployment, and then upgrade that deployment to version 4.4.4, the UI will show an error; the new **Load balancing** toggle will be turned on with associated settings displayed; and the UI will not allow you to save changes.

Workaround: First, toggle **Load balancing** off. This will return the UI to normal behavior. Then, if you want to use the Destination without load balancing, configure as usual and save. Or, if you prefer, update the **Webhook URLs** to include one or more valid endpoints you want to send to, and save the Destination with **Load balancing** enabled.

Fix: Version TBD.

2024-01-05 – v4.0.0-v4.4.3 – Windows Event Forwarder Source used up its allowed quota of connections [CRIBL-21965]

Problem: When the Windows Event Forwarder (WEF) Source could not process a request coming from a Windows client, it failed to correctly close the socket or send an error response to the client. Eventually, maximum allowable active connections could be reached, stopping the flow of data.

Workaround: For any affected WEF Source, set **Advanced Settings** > **Max active requests** to `0` to allow unlimited requests. Note that this will eventually cause other problems as the number of open-inactive sockets grows, meaning that the Worker nodes will eventually need to be restarted.

Fix: In Cribl Edge 4.4.4.

2023-12-21 – v.4.0.0-v4.5.1 – Splunk TCP Source logs include an unclear message about an unsupported payload [CRIBL-21845]

Problem: Because the Splunk TCP Source does not support ingesting compressed data via the Splunk S2S protocol, it cannot parse such payloads, and a correct error message in this situation would say “Could not parse payload. Turn compression off in upstream sender and try again.” Instead, the Source logs the unclear message “Failed to parse S2S payload”.

Workaround: None.

Fix: In Cribl Edge 4.6.

2023-12-13 v.4.0.0–4.4.3 – Sources incorrectly report failure [CRIBL-21690]

Problem: When using OpenTelemetry HTTP, the Source always listens on the default address (either `0.0.0.0` or `::`, or both, depending on the network configuration of the host), ignoring the option you configured in **General Settings** > **Address**.

Workaround: Unidentified.

Fix: In Cribl Edge 4.4.4.

2023-11-30 – v4.4.0-v.4.4.1 – Edge-only Sources are unavailable in standalone Edge instances [CRIBL-21443]

Problem: Edge-only Sources are disabled in standalone (not distributed) Edge instances.

Workaround: None.

Fix: In Cribl Edge 4.4.2.

2023-11-21 – v.4.4.0–4.4.3 – Azure Data Explorer Destination exposes irrelevant PQ settings in Cribl.Cloud [CRIBL-21335]

Problem: For Edge Nodes in Cribl.Cloud with Azure Data Explorer (ADX) Destinations: such Destinations' **Persistent Queue Settings** should expose only a **Clear Persistent Queue** button. They are incorrectly also exposing the PQ settings required for on-prem deployments.

Workaround: In Cribl.Cloud ADX Destinations, ignore all Persistent Queue settings except the **Clear Persistent Queue** button.

Fix: In Cribl Stream 4.4.4.

2023-11-16 – v4.4.0–4.4.3 – "Unsupported system page size" error when upgrading the Leader[CRIBL-21289]

Problem: Fedora users can experience an "unsupported system page size" error when upgrading the Leader. This is due to an issue with `jemalloc`.

Workaround: You can manually remove `jemalloc.so`. Docker users can use the following:

```
docker ... -e CRIBL_BEFORE_START_CMD_0="rm /opt/cribl/bin/libjemalloc.so" ...
```

Fix: In Cribl Edge 4.4.4.

2023-11-15 – v4.4.0 through Current – With Leader HA/Failover, creating diag bundles via CLI produces incorrect results [CRIBL-21258, CRIBL-21557]

Problem: For deployments configured for Leader high availability/failover, creating diagnostic bundles via the UI works normally but creating them via the CLI does not. In Cribl Edge v.4.4.0 and newer, the operation fails with a `not a git repository` error. In v.4.4.0 and older, the diag bundle created contains obsolete data. In both cases the underlying cause is that the process that creates the diag bundle cannot access the `CRIBL_CONF_DIR` environment variable.

Workaround: Use the UI to create the diag bundle. Or, if you use the CLI, preface the `diag` command with the needed environment variable, like this: `CRIBL_CONF_DIR=/<path_to_failover_volume> cribl diag create`. If you have an `instance.yml` config file, you can find the path to the failover volume in the `distributed` section.

Fix: Error message improvements were made as part of CRIBL-21258. Fix for CRIBL-21557 is TBD.

2023-11-09 – All Versions through Current – High-volume UDP data dropped in Cribl.Cloud [SAAS-4399]

Problem: Ingesting high rates of UDP events per second can cause Cribl.Cloud to drop some of the data. This limitation of the UDP protocol affects UDP-supporting Sources: Raw UDP, Metrics, SNMP Trap, and Syslog in UDP mode.

Workaround: To minimize the risk of data loss, deploy a hybrid Stream Worker Group, with Worker Nodes as close to the UDP senders as possible.

Cribl also recommends tuning the OS UDP buffer size.

Fix: Version TBD.

2023-11-09 – v.4.4.0–4.4.1 – Go version update required [CRIBL-21155]

Problem: The Go version requires an update.

Workaround: None.

Fix: In Cribl Edge 4.4.2.

2023-11-08 – v.4.4.0–4.4.1 – Azure Data Explorer Destination fails to validate database settings in Streaming mode [CRIBL-21123]

Problem: When Azure Data Explorer Destination is in **Streaming** mode and **Validate database settings** is turned on, database validation will fail unless **Ingestion service URI** has been set in **Batching** mode first.

Workaround: Switch to **Batching** mode, set the **Ingestion service URI** field, and save the Destination. Then switch back to **Streaming** mode and save the Destination again.

Fix: In Cribl Edge 4.4.2.

2023-11-07 – v.4.4.0–4.4.3 – Stream Groups' Members indicators always show zero count [CRIBL-21099, CRIBL-22072]

Problem: On a Stream Worker Group's **Manage** > **Overview** page, the **Members** counters show `0` in all Roles/Permissions, even when Members have been added.

Workaround: Navigate to **Group Settings** > **Members** for an accurate view of all Members and their access levels.

Fix: Fix for CRIBL-22072 in Cribl Edge 4.4.4. (CRIBL-21099 closed as a duplicate.)

2023-11-07 – v.4.4.0–4.4.1 – Worker Node generates an incomplete diagnostic bundle when Include git is enabled [CRIBL-21093]

Problem: When you teleport to a Worker Node and create a diagnostic bundle, if you leave **Include git** toggled on in the **Create & Export Diag Bundle** modal, the bundle that the system creates will be malformed and missing information, and the UI will produce a fatal error notification.

Workaround: When you create diagnostic bundles, turn **Include git** off.

Fix: In Cribl Edge 4.4.2.

2023-11-06 – v.4.4.0–4.4.1 – Azure Data Explorer Destination does not correctly turn off Validate database

settings [CRIBL-21060]

Problem: Once the Azure Data Explorer Destination has been saved with **Validate database settings** on, turning the setting off has no effect, and the Destination performs validation anyway.

Workaround: To ensure that **Validate database settings** is turned off, perform either of these two workarounds:

Option 1: Clone the Destination or create a new one, and turn **Validate database settings** off before saving.

Option 2: Turn off **Validate database settings**, then restart the Cribl Edge Edge Node.

Fix: In Cribl Edge 4.4.2.

2023-10-12 – v.4.3.0–4.4.1 – Rapid logging causes Edge Node memory spikes and prevents logging [CRIBL-20607]

Problem: A large number of logs written in a short time causes memory spikes in Edge Nodes and stops log output.

Workaround: Unidentified.

Fix: In Cribl Edge 4.4.2.

2023-10-05 – v.4.3.0 – Microsoft Windows Installer (MSI) does not create an Edge desktop shortcut [CRIBL-20504]

Problem: When you install Cribl Edge on Windows, the MSI doesn't create the expected desktop shortcut for Edge, even if you select **Create a shortcut for Cribl Edge on the desktop**.

Workaround: None.

Fix: In Cribl Edge 4.3.1.

2023-10-03 – v.4.3.0–4.4.1 – Upgrading to v.4.3.0 or later causes Amazon Kinesis Streams and WEF Sources to lose state [CRIBL-20444]

Problem:

After an upgrade to v.4.3.0, Edge Nodes for a Cribl Edge Amazon Kinesis Streams Source or Windows Event Forwarder (WEF) Source can lose state upon init. This can be especially problematic for customers with large streams and high data ingestion.

An Amazon Kinesis Streams Source that loses state will fail to read a data stream from the point where it most recently left off. Instead, the Source will start reading from the location configured by **Optional Settings** > **Shard iterator start**.

- If **Shard iterator start** is set to `Earliest Record` (the default), the Source can start too far behind in the stream to ever catch up.
- If **Shard iterator start** is set to `Latest Record`, the Source can start later than where it left off, causing some records to be skipped.

A Microsoft Windows Event Forwarder (WEF) Source (with bookmarks configured for the subscription) that loses state will fail to request that upstream clients send events starting at a bookmarked location. Instead, Cribl Edge will send no bookmark, prompting upstream clients to send **all** events that otherwise match the subscription. This can cause Cribl Edge to ingest duplicate events from WEF clients.

Workaround: If you have not yet upgraded to v.4.3.0 or later, consider delaying the upgrade until this issue is fixed.

Fix: In Cribl Edge 4.4.2.

2023-09-15 – v.4.3.0 – Scheduled (cron-based) jobs fail to load for some Sources and Collectors [CRIBL-20088]

Problem: Collection tasks for Office 365 Sources, the Splunk Search Source, the Prometheus Scraper Source, and scheduled Collectors fail to load when the Leader restarts, fails over, or upgrades. Collectors running jobs ad hoc are not affected.

Workaround: If you're using one of the listed Sources or use scheduled Collector jobs on-prem, do not upgrade to 4.3.0. If you are already on 4.3.0, downgrade to 4.2.2, or wait and upgrade to 4.3.1 when it's available.

You can also restart the config helper for on-prem installations using the command:
`pkill -f CONFIG_HELPER`

You must run this command each time the Leader fails over, or is restarted (ideally, not too often).

Fix: In Cribl Edge 4.3.1.

2023-09-14 – v.4.3.0 – Chain Function's link breaks when the chained Pipeline/Pack is resaved [CRIBL-20083]

Problem: After linking a Chain Function to a Pipeline or Pack, resaving the target Pipeline/Pack triggers `Invalid link` errors.

Workaround: 1. Copy the Pipeline containing the Chain Function to your clipboard, delete the original Pipeline, and then paste the copy to resolve the error. 2. Alternatively, on-prem admins can restart the config helper that manages the Fleet. (In the UI, navigate to **Settings** > **Global** > **System** > **Services** > **Processes**.)

Fix: In Cribl Edge 4.3.1.

2023-09-14 – v.4.3.0 – In Cribl.Cloud, Admin Members cannot modify Global Settings [CRIBL-20038]

Problem: After upgrading to 4.3.0, Members granted the Admin Permission cannot modify Global Settings.

Workaround: Members with the Owner Permission can still modify Global Settings.

Fix: In Cribl Edge 4.3.1.

2023-09-14 – v.4.3.0 – Edge Nodes cannot be upgraded in Leader-managed upgrades [CRIBL-20086]

Problem: In Stream and Edge, if you upgrade the Leader to 4.3.0, then use **Commit & Deploy** to push configs to Fleets before upgrading Edge Nodes to 4.3.0, you won't be able to upgrade the Edge Nodes from the Leader thereafter.

This issue affects only Leader-managed upgrades of Edge Nodes.

Workaround: If a Leader is on 4.3.0, you should upgrade all Edge Nodes to 4.3.0 before making any config changes or committing and deploying to the Nodes.

If you have already committed and deployed to Edge Nodes that are on a version prior to 4.3.0 (and the Leader is on 4.3.0), here are three workaround options:

- Revert and redeploy the last commit, then upgrade the Edge Nodes and deploy the most up-to-date changes, **OR**
- Upgrade the Edge Nodes via command line or script, **OR**
- Downgrade the Leader to 4.2.2 and then upgrade to 4.3.1 when it is available.

If you run into issues, contact support@cribl.io for resolution assistance.

Fix: In Cribl Edge 4.3.1.

2023-09-13 – v.4.2.1 through Current – Edge upgrades occasionally fail when Fleets contain a large number of Nodes [CRIBL-20067]

Problem: Leader-managed upgrades occasionally fail when a Fleet contains a large number of Nodes.

Workaround: Restart the upgrade operation to upgrade additional Nodes.

Fix: Version TBD.

2023-09-13 – v.4.3.0 – Any Code Function before another Function breaks the Data Preview OUT tab display [CRIBL-20040]

Problem: Any Code Function inserted before the end of a Pipeline prevents the Data Preview OUT tab from rendering properly. This occurs even if the Code Function contains only a single comment, or is blank. Adding a Code Function at the end of a Pipeline does not cause an issue. This is a Data Preview UI issue only – if you capture events to a Destination, Cribl Edge processes them as expected.

Workaround: None.

Fix: In Cribl Edge 4.3.1.

2023-08-31 – v.4.2.2 – When in XML format, the Windows Event Logs Source doesn't accept log names that contain spaces [CRIBL-19818]

Problem: When the Windows Event Logs Source is set to XML format, any configured log names that contain spaces (e.g., `Windows Powershell`) cause an error and the log can't be read.

Workaround: None.

Fix: In Cribl Edge 4.3.

2023-08-28 – v.4.2.2 through 4.3.0 – PQ doesn't drain for the Event Hubs Destination when Acknowledgements is set to All or Leader [CRIBL-19756]

Problem: When the `Acknowledgements` setting is set to `All` or `Leader`, Kafka-based Destinations (especially Azure Event Hubs) can fail to drain the persistent queue (PQ) and the logs are filled with `Attempting to send faster than the downstream can receive – consider checking the network connection and broker health` errors.

Workaround: You can set a limit using the `Persistent Queue Drain rate limit (EPS)` setting (for example, set it to `500`) to get the PQ to drain when the datagen is turned off. However, if the rate of data input continues to exceed the rate at which the Destination can send events downstream (which is slowed by the `Acknowledgements` setting), the PQ will continue to build. Alternatively, you can set the `Acknowledgements` setting to `None`, which significantly increases the rate at which data can be sent to Event Hubs.

Fix: In Cribl Edge 4.3.1.

2023-08-24 – v.4.2.0 through 4.2.2 – Deleting a Source via QuickConnect doesn't refresh the view [CRIBL-19709]

Problem: When using QuickConnect, if you delete a Source by opening its drawer and clicking `Delete Source`, the Source is deleted. However, it will continue to show on the page until you manually refresh the page.

Workaround: Manually refresh the page after deleting a Source.

Fix: In Cribl Edge 4.3.0.

2023-08-23 - v.4.0.0 through 4.4.4 - Packs installed in Fleets are not visible to Subfleets. [CRIBL-19676]

Problem: When a Fleet has Subfleets, and you install a Pack in the parent Fleet, none of the Subfleet's Pipeline drop-downs make the Pack available. This is true for Routes, pre-processing Pipelines in Sources, and post-processing Pipelines in Destinations.

Workaround: Install the Pack both in the parent Fleet, and in any Subfleets that need to use the Pack.

Fix: In Cribl Edge 4.5.

2023-08-23 – v.4.0–4.2.x – Inherited Packs in Fleets lose their configured Routes [CRIBL-19700]

Problem: Routes in a Pack at the Fleet level revert to default Routes when inherited by a SubFleet.

Workaround: None.

Fix: Couldn't reproduce the error.

2023-08-16 – v.4.2.2 – When adding or editing a Route, typeahead misbehaves in the Filter field [CRIBL-19571]

Problem: Entering text in the **Filter** field of a Route is difficult in Cribl Edge 4.2.2 because the typeahead function behaves erratically, sometimes moving the cursor around or inserting text fragments.

Workaround: 1. Assemble the filter expression in a different Cribl expression field, or in any text editor, then copy and paste it here. 2. Alternatively, click the Manage as JSON button at the **Data Routes** table's upper right. Then, in the JSON editor, build the expression as the `"filter":` key's value.

Fix: In Cribl Edge 4.3.

2023-08-02 - v.4.2.1 through Current - Healthy Destinations spuriously report `Blocked Status` [CRIBL-19322]

Problem: Healthy TCP-based, Kafka-based, Splunk, and Metrics Destinations can spuriously report `Blocked Status` in the **Charts** tab, sometimes also logging the blocked status inaccurately.

Workaround: None.

Fix: Version TBD.

2023-07-26 – v.4.2.0–4.2.1 – File Monitor Source causes memory leaks and Edge Node crashes [CRIBL-19154]

Problem: The File Monitor Source, when running in Manual Discovery mode, leaks memory and a file descriptor each time it rediscovers a file that it has already collected. The leaked resources increase with every polling interval, for every rediscovered file, eventually causing Edge Nodes to crash. This does not affect Auto Discovery mode, and does not affect deployments without an active File Monitor Source.

Workaround: File Monitor users should bypass (or roll back from) versions 4.2.0–4.2.1.

Fix: In Cribl Edge 4.2.2.

2023-07-20 – v.4.2.0–4.2.1 – AppScope Source might reference a nonexistent config [CRIBL-19044]

Problem: The 'sample_config' stock configuration was removed in v.4.2.0. This missing config causes the Appscope Source to disappear from the UI, and changes to other Sources also fail with validation errors.

Workaround: For on-premises/customer-managed deployments, clone one of the existing AppScope configurations and rename it to `sample_config`. For Cribl.Cloud customers, there is no workaround yet.

Fix: In Cribl Edge 4.2.2.

2023-07-19 – v.4.2.2 – Group Information for Cribl.Cloud Worker Groups not showing on summary page for SSO users [CRIBL-19107]

Problem: On Cribl.Cloud Organizations, when you open a Stream Group's **Manage** > **Overview** page, the **Group Information** section is not populated for users imported using SSO from external identity providers. (This section populates correctly for users/Members configured natively within your Cribl Organization.)

Workaround: None.

Fix: Couldn't reproduce the error.

2023-07-19 – v.4.2.0–4.2.1 – Users with the Admin Permission cannot view Notifications [CRIBL-19015]

Problem: Users with the **Admin** Permission receive a `You do not have sufficient permissions to access this resource` message when attempting to view Notifications. Users with this Permission should be able to access Notifications.

Workaround: None. Organization **Owners** can access these notifications.

Fix: In Cribl Edge 4.2.2.

2023-07-19 – v.4.2.0–4.2.1 – Users with the Admin Permission cannot bootstrap a Worker due to a missing auth token [CRIBL-19014]

Problem: Users with the **Admin** Permission can't add a Worker using **Add / Update Worker Node**, because the modal doesn't have the **Leader hostname/IP** or the **Auth token** fields populated. You can enter the hostname/IP into its field, but not the auth token. Users with the **Admin** Permission should see the auth token.

Workaround: Organization **Owners** can perform this action and share the auth token.

Fix: In Cribl Edge 4.2.2.

2023-07-19 – v.4.2.0 – Credential encryption errors preventing user access to Cribl Edge Sources and Destinations [SAAS-4823]

Problem: A critical regression caused Fleets to become unresponsive, preventing access to Sources and Destinations. This issue was caused by incorrectly encrypted credentials.

Workaround: Upgrade to Cribl Edge 4.2.1.

Fix: In Cribl Edge 4.2.1.

2023-07-18 – v.4.2.0 through Current – Migrated Local Users appear in the Members UI with “No Access” [CRIBL-18973]

Problem: Local Users are incorrectly shown in the **Settings** > **Members** UI with `No Access`. However, their Roles still function as originally configured, and still display correctly at **Settings** > **Global Settings** > **Access Management** > **Local Users**.

Workaround: Rely on the **Local Users** UI.

Fix: Version TBD.

2023-07-18 – v.4.2.0–4.4.4 – Product-level Editor Permissions do not allow a commit/deploy action on the Worker Groups/Fleets page [CRIBL-18970, CRIBL-20113]

Problem: Users with product-level Edge **Editor** or Stream **Editor** Permissions are unable to commit and deploy changes made to Worker Groups or Fleets from the Group or Fleet page. However, if the user navigates to the Group or Fleet where the change was saved, the **Commit and Deploy** button is available.

Workaround: Commit and deploy from the Worker or Group where you saved the change.

Fix: In Cribl Edge 4.5. CRIBL-18970 is closed and expanded to CRIBL-20113.

2023-07-18 – v.4.2.0–4.2.1 – Users with the Edge Editor Permission cannot delete secrets under Fleet Settings [CRIBL-18969]

Problem: Users with the Edge **Editor** Permission can create secrets under **Fleet** > **Fleet Settings** > **Secrets** but they are unable to delete the secret once it is created. Edge **Editors** should have this capability.

Workaround: Assign the Edge **Admin** Permission to users that need to be able to delete secrets.

Fix: In Cribl Edge 4.2.2.

2023-07-17 – v.4.2.0–4.2.1 – Edge Node Settings not visible when teleporting [CRIBL-18963]

Problem: When you teleport into an Edge Node, you can't view **Node Settings** in the UI.

Workaround: None.

Fix: In Cribl Edge 4.2.2.

2023-07-17 – v.4.2.0–4.2.1 – Users with the Edge Editor Permission cannot delete Sources, Destinations, or Pipelines [CRIBL-18951]

Problem: If you assign a user the Edge **Editor** permission at the product-level, the user will not be able to delete Sources, Destinations, or Pipelines. Edge **Editors** should have this capability.

Workaround: Assign the Edge **Admin** Permission to users that need to be able to delete these items.

Fix: In Cribl Edge 4.2.2.

2023-07-17 – v.4.2.0–4.2.1 – Change from Edge User to Edge Editor can fail to take effect [CRIBL-18950]

Problem: If you assign the **Editor** Permission to an Edge **User** at the Fleet-level and then later assign the Edge **Editor** Permission to the same User at the product-level (Edge **User** now becomes an Edge **Editor**), the **Editor** Permission at the product-level can fail to take effect. The user will have **Editor** access to the one Fleet at the Fleet-level but they will still show as an Edge User and will not have **Editor** Permissions at the product-level.

Workaround: None.

Fix: In Cribl Edge 4.2.2.

2023-07-17 – v.4.2.0–4.2.1 – Product logins for users with product-level Read Only and Editor Permissions can result in misleading log entries [CRIBL-18948]

Problem: When users that are assigned product-level **Read Only** or **Editor** Permissions log in to a product, the browser incorrectly makes requests for resources that users at this level are not allowed to access. While these requests are correctly denied and there is no risk of compromising security, Cribl admins may see misleading log entries indicating that these users tried to access something they are not allowed to access.

Workaround: None.

Fix: In Cribl Edge 4.2.2.

2023-07-16 – v.4.2.0–4.4.3 – After upgrading, Monitoring dashboards are not visible for users with `GroupRead` and `GroupCollect` Policies [CRIBL-18925]

Problem: Users that were assigned `GroupRead` and `GroupCollect` Policies in Cribl Edge 4.1.x will no longer be able to access the **Monitoring** tab and dashboard after upgrading to Cribl Edge 4.2.

Workaround: For `GroupRead`, migrate the user to **Read Only** Permission at the Group-level in Members and Permissions. Assign the Member as Stream **User** then set their Permissions for the group as **Read Only**. For `GroupCollect`, there is no specific workaround but users can be assigned the `GroupEdit` Policy if they need to view Monitoring dashboards.

Fix: In Cribl Edge 4.4.4.

2023-07-15 – v.4.2.0–4.2.1 – Edge users with the Fleet-level Editor Permissions cannot use manual file discovery [CRIBL-18923]

Problem: When using manual file discovery in Cribl Edge, users that have been assigned the **Editor** Permission at the Group/Fleet-level will not see any results.

Workaround: You can assign the **Admin** Permission for users that require the manual file discovery feature.

Fix: In Cribl Edge 4.2.2.

2023-07-13 – v.4.2.0–4.2.1 – Stream Users are able to access the Monitoring page [CRIBL-18848]

Problem: Stream **Users** with no Group-level Permissions are able to access the **Monitoring** page. The **Monitoring** menu item is hidden, but the page can be access manually using a URL.

Workaround: None.

Fix: Not observed as of Cribl Edge 4.2.2.

2023-07-13 – All Versions through Current – When configuring an S3 Collector, JavaScript expressions break the "Auto-populate from" option for the Path field [CRIBL-18844]

Problem: When automatically populating an S3 Collector from an S3 Destination, the Collector **Path** field won't resolve JavaScript expressions, even if the expression was valid on the Destination that the field was populated from. The S3 Collector path is treated like a literal string and the software fails to warn you that the path is invalid.

Workaround: None.

Fix: In Cribl Edge 4.3.

2023-07-12 – All Versions through Current – Recent Actions endpoint returns results for certain unauthorized users [CRIBL-18818]

Problem: The Recent Actions endpoint `GET /api/v1/ui/recentActions` returns more recent actions than intended for non-admin users.

Workaround: None.

Fix: In Cribl Stream 4.2.2.

2023-07-12 – v.4.2.0–4.2.1 – No Access to Subfleets for the Edge User Permission [CRIBL-18784]

Problem: Regardless of what Permissions are granted at the Fleet-level, Edge members assigned to the **User** Permission will not have access to Subfleets.

Workaround: At the (Edge) product level, you must grant your members **Read Only**, **Editor**, or **Admin** Permissions so that they can access Subfleets.

Fix: In Cribl Edge 4.2.2.

2023-07-07 – v.4.1.3-4.2.1 – When configuring an HTTP Source, enabling Source PQ changes the inputId in events [CRIBL-18668]

Problem: When Source-side Persistent Queueing is enabled on an HTTP Source, the trailing colon is dropped from `__inputId`. This will break Routes and filters that are based on the Input ID if it is copied from the configuration page. Live captures will also stop working until you modify them.

Workaround: None.

Fix: In Cribl Edge 4.3.

2023-07-06 – v.4.2.0–4.2.1 – Default Pack is not visible to Project Editor when a Project is shared with them [CRIBL-18639]

Problem: When a Cribl admin sets up a Project, a default Pack is included. If the Admin then shares that Project with a Project editor, the editor will not see the default Pack.

Workaround: In the Project view, the Project editor can add any Pack that has been made available by the admin, including the default Pack.

Fix: In Cribl Edge 4.3.

2023-07-03 – v.4.2.0–4.3.0 – Users with Group-level Read Only Permissions can interact with logging level menus [CRIBL-18556]

Problem: Users with the **Read Only** Permission at the Group-level can access the **Add Channel** and **Delete** buttons under **Group** > **Manage** > **Group Settings** > **Logging** > **Levels**. However, they are unable to save changes. When performing these actions, **Read Only** users will see a **Forbidden** message displayed in the UI and no changes will be saved. The setting should not be accessible to users with this Permission.

Workaround: None.

Fix: In Cribl Edge 4.3.1.

2023-06-29 – All Versions through v.4.1.3 – (Linux) System Metrics Source does not emit “Per interface” metrics when

set to “All” [CRIBL-18473]

Problem: In the (Linux) **System Metrics** Source > **Host Metrics** tab, selecting **All** does not emit **Per Interface** metrics.

Workaround: In the **Host Metrics** tab, select **Custom** > **Network** > **Custom** then toggle **Per interface metrics** to `Yes`.

Fix: In Cribl Edge 4.2.

2023-06-28 – v.4.2.0–4.2.1 – Project editor can’t delete Pipelines via Options menu [CRIBL-18447]

Problem: In a Project’s **Pipelines** modal, when a Project editor opens a Pipeline’s ••• (Options) menu, the **Delete** option is unavailable. Cribl Edge displays a spurious error about insufficient permissions.

Workaround: Use the list’s check boxes to select Pipelines, then click **Delete Selected Pipelines**.

Fix: In Cribl Edge 4.2.2.

2023-06-26 – All Versions through v.4.1.3 – High CPU usage in File Monitor Source’s Manual mode [CRIBL-18400]

Problem: In the **File Monitor** Source and the **Explore** > **Files** tab UI, the **Manual** mode does not honor the **Max depth** setting. The API process consumes high CPU resources because the discovery logic (used in the **File Monitor** Source and **Files** tab) recurses in the directory tree.

Fix: In Cribl Edge 4.2.

2023-06-26 – All Versions through Current – Fleet Upgrade Errors [CRIBL-18374]

Problem: A Fleet’s upgrade from a Leader can result in errors when the Edge Nodes use different `host/port/tls` settings than Worker Nodes.

Workaround: The Edge Nodes and Workers must connect to the leader using the `host/port/tls` connection details. If this is not possible, upgrade Edge Nodes separately using the boostrap script.

Fix: Version TBD.

2023-06-21 – v.3.5.2 through Current – Datadog Agent Source fails to receive metrics [CRIBL-18281]

Problem: When a Datadog Agent using Datadog API v2 sends metrics, Datadog Agent Source fails to receive them, because the Source uses Datadog API v1. Datadog Agent logs will show `API Key invalid, dropping transaction` errors. Debug logging will likely produce `Dropping request because of unallowed path message` errors with `statusCode: 403`, which Datadog Agent interprets as "invalid API key."

Workaround: In the `datadog.yaml` file, ensure that `use_v2_api.series` is set to `false`. Cribl Edge's Datadog Agent Source will then receive metrics normally, since both the Source and the Datadog Agent will be using Datadog API v1, which Datadog still supports.

Fix: Version TBD.

2023-06-21 – v.3.5.2–4.1.3 – Using the Rename Function gives unexpected results due to skipped internal fields [CRIBL-18285]

Problem: When internal fields are present in the `Rename fields` list, the Rename Function may incorrectly assign field keys or values.

Workaround: This function is not intended to operate on internal fields. Avoid this operation.

Fix: In Cribl Edge 4.2.

2023-06-20 – All versions through 4.2.2 – HTTP Destinations sometimes send oversized payloads [CRIBL-18218]

Problem: HTTP-based Destinations do not strictly enforce the **Max Body Size** limit in all cases. This can cause payloads to be sent that exceed the maximum size some downstream receivers can accept.

Workaround: Experiment with lower **Max Body Size** values until downstream receivers are reliably accepting all events.

Fix: In Cribl Edge 4.3.

2023-06-15 – v.4.1.3 – Amazon CloudWatch Destinations log error while flushing events older than 24 hours [CRIBL-18184]

Problem: Amazon CloudWatch Destination doesn't filter batches of events to ensure all events in the batch are within 24 hours of each other as required by the API, resulting in the batch being rejected by

Cloudwatch.

Workaround: Add a Post-Processing Pipeline with a Drop Function using a Filter Expression of `_time<(Date.now() / 1000) - 86400`.

Fix: In Cribl Edge 4.2.

2023-06-14 – v.4.1.2–4.1.3 – The Office 365 Activity Source misses events [CRIBL-18164]

Problem: The Office 365 Activity Source misses events due to the current collection system. For services such as Exchange, SharePoint, OneDrive, and Teams, Microsoft indicates that audit record availability is typically 60 to 90 minutes after an event occurs. Our current collector methodology does not account for this availability window. Instead, the Source completes runs as scheduled and collects only events it finds during the configured time range, which can lead to missed events.

Workaround: None.

Fix: In Cribl Edge 4.2.

2023-06-08 – v.4.1.2 – AES-256-GCM security vulnerability [CRIBL-18038]

Problem: The AES-256-GCM encryption option introduced in v.4.1.2 included a security vulnerability.

Workaround: Do not use this option in v.4.1.2.

Fix: In Cribl Edge 4.1.3.

2023-06-01 – v.4.0.0–4.1.3 – GroupEdit is not able to commit changes [CRIBL-17939]

Problem: Users having the `GroupEdit` policy are able to make changes (e.g. create new Sources) but are unable to **Commit** their changes. These users should be able to **Commit**, but not **Deploy**, changes.

Workaround: Apply the `GroupFull` policy for users that need to be able to **Commit** changes. These users will also have the ability to deploy a Worker Group or Fleet.

Fix: In Cribl Edge 4.2.

2023-06-01 – v.4.1.0–4.1.2 – Kubernetes Logs Event Breaker Incorrectly Breaks Events [CRIBL-17907]

Problem: The **Kubernetes Logs** Event Breaker incorrectly breaks events with out-of-order timestamps.

Workaround: In the Event Breaker, move the logic that adjusts `_raw` into a **Pre-Processing** Pipeline.

Fix: In Cribl Edge 4.1.3.

2023-05-27 – v.4.1.1–4.1.2 – Perf bug: RPC consuming excessive CPU compiling expressions [CRIBL-17779]

Problem: In 4.1.1, to improve product security, we changed how we manage our RPC traffic from Leaders and Workers. This change can have a negative impact in larger environments with many Worker Processes. Following a 4.1.1 or 4.1.2 upgrade, the Leader can become non-responsive to UI requests due to a steady 100% CPU utilization. When this happens, you cannot manage or monitor Cribl Edge environments.

Workaround: Roll back to a previous stable version, such as 4.1.0.

Fix: In Cribl Edge 4.1.3.

2023-05-26 v.4.1.2–4.1.3 – Edge Nodes don't update from Cribl Edge Leader [CRIBL-17792]

Problem: In Cribl Edge 4.1.2–4.1.3, if you define the `baseURL` key on a Leader, Edge Nodes don't get the latest config version from the Leader. The log may also contain `checksum mismatch` warnings.

Workaround: Clear **URL base path** in **Settings** > **General Settings** > **Advanced** or specify an empty `baseURL` key in `cribl.yml`. If you need to define a base URL, do not upgrade to the affected versions.

Fix: In Cribl Edge 4.2.

2023-05-19 – v.4.1.2 – CrowdStrike Destination's "LogScale endpoint" field hidden from UI [CRIBL-17715]

Problem: The CrowdStrike Falcon LogScale Destination's **LogScale endpoint** field is hidden from the 4.1.2 configuration modal, but can be restored.

Workaround: Follow these steps, in sequence, for each LogScale Destination.
1. Populate the LogScale config modal, and save once.
2. Reopen the modal, then select **Manage as JSON**.
3. Change the `"loadbalanced":` key's value from `true` to `false`.
4. Copy this default nested element: `"url": "https://cloud.us.humio.com/api/v1/ingest/hec",`
5. Paste it above the whole `"urls"` element (typically at line 21).

6. Click **OK** to restore the visual UI, with the **LogScale endpoint** field now visible.
7. Enter your actual endpoint URL, and save the config.

Fix: In Cribl Edge 4.1.3.

2023-05-16 – v.4.1.1–4.1.2 – Expanding Status of Output Router Destination triggers error [CRIBL-17632]

Problem: When you attempt to expand a node on the **Status** tab for an Output Router Destination, the page crashes and the error `Cannot convert undefined or null to object` appears.

Fix: In Cribl Edge 4.1.3.

2023-05-12 – v.4.0–4.1.3 – Windows Event Forwarder Source causes duplicate events when experiencing backpressure [CRIBL-17614]

Problem: In certain cases where the Windows Event Forwarder Source experiences backpressure from a downstream Pipeline or Destination, it will return an incorrect error message to the client that sent the events. This causes that client to re-send the same events and results in duplicate events ingested in Stream.

Workaround: In cases where the cause of the backpressure can be resolved, this issue is mitigated. However if the backpressure cannot be completely removed, there is no additional workaround.

Fix: In Cribl Edge 4.2.

2023-05-12 – v.4.0–4.1.2 – The Kubernetes Logs Source drops events [CRIBL-17602]

Problem: Events from a running container will stall when the underlying container runtime rotates the log file. Any further rotation of log files will result in data loss.

Workaround: Restart the Kubernetes Logs Source to reconnect.

Fix: In Cribl Edge 4.1.3.

2023-05-05 – v.4.1.1 – S3 ingestion slows after upgrade [CRIBL-17317]

Problem: After upgrading to Cribl Edge 4.1.1, ingesting using any `JSON` Array Event Breaker (such as AWS CloudTrail) will slow or stop due to high CPU usage.

Workaround: Roll back to a previous stable version, such as 4.1.0.

Fix: In Cribl Edge 4.1.2.

2023-04-27 – v.4.1.1 – Edge Nodes integrated with `systemd/initd` can fail upon reconfig [CRIBL-17264]

Problem: Where 4.1.1 on-prem or hybrid Edge Nodes (on Linux) are integrated with `systemd/initd`, deploying new config bundles to those Edge Nodes can stop the Edge Nodes from processing data. The root cause is an incorrect backup step, which (depending on permissions) can also consume large amounts of disk space on their hosts.

Precondition: Cribl-managed Cloud instances are unaffected. This problem has been observed only if at least one of the following directories is populated, or exists with restricted permissions:

```
/data/
/default/cribl/
/default/edge/
/local/cribl/
/local/edge/
/default/<any-installed-Pack-name>/
/local/<any-installed-Pack-name>/
```

Workaround: Skip v.4.1.1 or roll back to your last stable version. To run v.4.1.1, this workaround is available on systemd's `Service` section: 1. In the script `/etc/systemd/system/cribl.service`, add the line: `WorkingDirectory=/opt/cribl`, (This is the default path – specify your own path equivalent to `$CRIBL_HOME`). 2. Then reload the systemd daemon: `systemctl daemon-reload`. 3. Then restart the Cribl Stream instance using `systemctl restart <service name>`.

Fix: In Cribl Edge 4.1.2.

2023-04-19 v.4.4.0–4.5.0 – Special characters in Source or Destination auth tokens cause authentication failure [CRIBL-17047]

Problem: When you include special characters in a Source or Destination auth token, those characters may cause an authentication failure.

Workaround: Use only alphanumeric characters or the underscore (`_`) in a Source or Destination auth token. Do not use any of the following characters: `<`, `>`, `"`, `` ` ``, `\r`, `\n`, `\t`, `{`, `}`, `|`, `\`, `^`, or `'`. Alternatively, upgrade to Cribl Edge 4.5.1.

Fix: In Cribl Edge 4.5.1.

2023-04-17 – All Versions through v4.2.2 – Kafka-based Destinations fail to report backpressure properly [CRIBL-16944]

Problem: With a downstream receiver exerting backpressure, Kafka-based Destinations do not report backpressure promptly, even though they effectively exert backpressure. As a result, upstream data might stall for several minutes before appropriate errors surface in the Destination's health status, and PQ (if configured) will not engage promptly.

Workaround: Decreasing the value for the `Request timeout (ms)` setting within the `Advanced Settings` section of the destination configuration might lead to the Destination reporting backpressure more promptly.

Fix: In Cribl Edge 4.2.2.

2023-04-14 – v.2.2–4.1.x – Requests from Office 365 Message Trace Source failed intermittently [CRIBL-16929]

Problem: Microsoft has fixed this Office 365 Message Trace API problem with this patch (Microsoft login required). Before this patch, requests failed intermittently, often with `Non-whitespace` errors.

Fix: Fixed by Microsoft (not a Cribl problem).

2023-04-13 – v.4.1.1 – Monitoring incorrectly shows metrics for disconnected Subscriptions [CRIBL-16926]

Problem: Selecting **Monitoring** > **Data** > **Subscriptions** will falsely show statistics even for Subscriptions that are not connected to a Destination, and therefore have no data flow.

Workaround: Ignore these Monitoring statistics.

Fix: In Cribl Edge 4.1.2.

2023-04-03 – v.4.1.0 – Windows Event Logs Source should have a configurable Max Event Size [CRIBL-16549]

Problem: Events aren't broken properly when collecting logs from the PowerShell event log. The Source contains a Max Event Limit of `51200`, which is a hard-coded breaker config. This limit should be configurable.

Fix: In Cribl Edge 4.1.1.

2023-03-31 – v.4.1.0-4.1.1 – Cribl.Cloud API Credential's Roles aren't honored on tokens [SAAS-3681]

Problem: API Bearer tokens obtained through the Cribl.Cloud portal are all granted the `Admin` Role, regardless of the scope specified on the parent Credential.

Workaround: Use the pre-4.0 workaround to obtain Bearer tokens from the in-app API Reference.

Fix: In Cribl Edge 4.1.2.

2023-03-30 v.4.1.0 – Splunk Load Balanced Destination incorrectly re-creates all connections during DNS resolution [CRIBL-16494]

Problem: When **Minimize in-flight data loss** is enabled, the Splunk Load Balanced Destination re-creates all outbound connections during DNS resolution. This will cause the Destination to report a blocked status until the outbound connections refresh. If persistent queues (PQ) are enabled, PQ will engage while the Destination is blocked.

Workaround: Select the Destination's **Manage as JSON** option, to add the key-value pair: `"maxFailedHealthChecks": 1`.

Fix: In Cribl Edge 4.1.1.

2023-03-28 – v.4.1.0 – Add Kubernetes and Docker modals in the UI aren't checking the `CRIBL_BOOTSTRAP_HOST` environment variable [CRIBL-16432]

Problem: The Add Windows and Linux modals check for a bootstrap hostname to see if a user has configured a host override with the `CRIBL_BOOTSTRAP_HOST` environment variable. The Add Kubernetes and Docker modals aren't checking the environment variable as expected.

Fix: In Cribl Edge 4.1.1.

2023-03-23 – v.4.1.0 – Leader's config deployments to pre-4.1 Workers silently fail [CRIBL-16339]

Problem: A 4.1.x Leader requires Workers to also be upgraded to – v.4.1.x. If you attempt to deploy configs to Workers running earlier versions, the intended upgrade prompt might not appear. In this case, the deploy

will silently hang.

Workaround: If you haven't enabled automatic upgrades of on-prem or hybrid Workers, upgrade all Fleets to 4.1.x before you deploy to them.

Fix: In Cribl Edge 4.1.1.

2023-03-22 – v.4.1.0-4.1.1 – Cribl Stream Database Collector cannot connect with SQL Server using AD authentication [CRIBL-16304]

Problem: On the indicated versions, the Database Collector cannot connect with SQL Server using Active Directory (AD) authentication. (Only local auth works.)

Workaround: Reverting to Cribl Stream v.4.0.4 restores the Database Collector's ability to connect using AD authentication.

Fix: In Cribl Stream 4.1.2.

2023-03-21 – v.4.1.0-4.3.1 – Unable to bind `CRIBL_DIST_MASTER_URL` to an IPv6 address [CRIBL-16284]

Problem: Due to an issue with the `CRIBL_DIST_MASTER_URL` environment variable, Cribl will not bind to a specified IPv6 address.

Workaround: Manually update the settings in the two relevant configuration files, `instance.yml` and `cribl.yml`, for RPC and UI communication respectively.

For the UI, manually define the host setting in the `cribl.yml` file:

```
api:
  host: "::"
```

To configure the instance as a Leader node and listen on all IPv6 and IPv4 addresses, manually configure the `instance.yml` file:

```
distributed:
  mode: master
  master:
    host: "::"
    port: 4200
    tls:
      disabled: true
    authToken: criblmaster
```

Fix: In Cribl Edge 4.4.

2023-03-20 – v.3.5.0–4.1.0 – System Activity drawer on Windows Edge Nodes displays no data [CRIBL-16194]

Problem: When you navigate to a Fleet's **List View** and click a row containing a Windows Edge Node, the System Activity drawer displays no data.

Fix: In Cribl Edge 4.1.1.

2023-03-20 – v.4.1.0-4.2.1 – If a deployment fails, Workers will not automatically revert to the previous version [CRIBL-16203]

Problem: When a deployment fails, Workers cannot revert `default/cribl` to a previous version because that directory is no longer backed up. The Worker will enter a broken state if you provide an invalid deploy bundle, because it cannot revert to the last valid state.

Workaround: To resolve any broken Workers, you must deploy a valid configuration. The Worker will resume working once it picks up the new configuration.

Fix: In Cribl Edge 4.3.1. Closed as a duplicate of CRIBL-15467 which was fixed in 4.1.1.

2023-03-15 – v.4.1.0 – QuickConnect Pipeline Settings button fails [CRIBL-16142]

Problem: When adding a new Pipeline using QuickConnect, clicking the gear ( ) button fails to open Pipeline Settings.

Workaround: After saving the Pipeline in QuickConnect, use **Manage** > **Processing** > **Pipelines** to select your new Pipeline. Click the gear button here to access Pipeline Settings as expected.

Fix: In Cribl Edge 4.1.1.

2023-03-14 v.4.0.0–4.1.0 – Duplicate data with Event Hubs [CRIBL-16102]

Problem: When **Minimize Duplicates** is enabled, the Azure Event Hubs Source assigns partitions to multiple Edge Nodes in the same Fleet, which generates duplicate data.

Workaround: Toggle **Minimize Duplicates** to `No` in **Azure** > **Event Hubs** > **Advanced Settings**.

Fix: In Cribl Edge 4.1.1.

2023-03-01 – v.4.0.4 – Kafka-based Sources are omitting the `_time` field [CRIBL-15696]

Problem: Kafka-based Sources (Kafka, Azure Event Hubs, Confluent Cloud) are not sending the `_time` field.

Workaround: If Cribl Edge is not retrieving the `_time` field, then in the affected Sources' config modals, use the **Processing Settings** > **Fields** tab to re-create `_time`. This new `_time` field will add the current (ingestion) time instead of the message time.

Fix: In Cribl Edge 4.1.

2023-02-24 – v.4.1.0-4.2.1 – The Chain Function has a 10% impact on performance [CRIBL-15538]

Problem: Using the Chain Function to chain data processing from one Pipeline or Pack to another degrades performance by about 10%, compared to running the original Pipeline or Pack directly.

Fix: In Cribl Edge 4.2.2.

2023-02-23 – v.4.0.x – DNS Resolution reconnects Cribl TCP connections every time [CRIBL-15363]

Problem: All Cribl TCP connections will reconnect every time DNS Resolution occurs, even when reconnection is not necessary.

Fix: In Cribl Edge 4.1.

2023-02-23 – v.4.0.4–4.1.2 – Enable Automatic Upgrades deletes remote repo's Git Settings [CRIBL-15502]

Problem: Enabling the Leader's **Upgrade** > **Enable Automatic Upgrade** setting deletes the corresponding remote repo's stored **Git Settings**.

Workaround: Reconfigure your repo at **Git Settings** > **Remote**.

Fix: In Cribl Edge 4.1.3.

2023-02-22 – v.4.0.4–4.1.0 – Sending data to an inactive Kafka, Confluent Cloud, or Azure Event Hubs Destination triggers an endless series of log errors [CRIBL-15455]

Problem: If you accidentally make a Kafka, Confluent Cloud, or Azure Event Hubs Destination inactive by putting a mismatched value in the **Advanced Settings** > **Environment** setting, Cribl Edge will send an endless series of errors to the log file.

Workaround: Correct the **Environment** value or leave it empty.

Fix: In Cribl Edge 4.1.1.

2023-02-20 – v.4.0.x–4.1.0 – Logging-level changes do not take effect for services [CRIBL-15365]

Problem: Changing logging levels in the Leader's Settings has no effect on the logs for the Connections, Lease Renewal, Metrics, or Notifications services.

Fix: In Cribl Edge 4.1.1.

2023-02-20 – v.4.0.3–4.1.0 – On-prem Edge Nodes automatically upgrade, ignoring Leader settings [CRIBL-15367]

Problem: Upgrading a Leader Node causes its on-prem Edge Nodes to automatically upgrade, even when the Leader's **Enable automatic upgrades** option is set to `No` (the default). This problem does not affect Cribl-managed Cribl.Cloud Edge Nodes.

Workaround: 1. Toggle **Enable automatic upgrades** to `Yes` and save the configuration; then slide it back to `No` and save again. Or: 2. Edit `local/cribl/cribl.yml` to explicitly set `upgradeSettings.disableAutomaticUpgrade` to `true`.

Fix: In Cribl Edge 4.1.1.

2023-02-14 – v.4.0.0–4.0.4 – Pipeline is deselected in QuickConnect after modification [CRIBL-15246]

Problem: If you add a Pipeline in the QuickConnect UI, then modify it in the **Edit Pipeline** modal, the Pipeline is deselected in the **Add Pipeline to Connection** modal. If you then close the **Add Pipeline to Connection** modal, the Pipeline will disappear from the Route.

Workaround: Select the modified Pipeline in the **Add Pipeline to Connection** modal before closing it.

Fix: In Cribl Edge 4.1.

2023-02-14 – v.3.5.0 through Current – Cannot clear Messages drawer while in GitOps Push mode [CRIBL-15239]

Problem: When in GitOps `Push` mode, you cannot clear messages from the **Messages** drawer. Cribl Edge returns a `Forbidden` error.

Fix: Version TBD.

2023-02-06 – v.4.0.x–4.1.0 – Event Breaker timestamp extraction fails after configured time [CRIBL-15107]

Problem: Event Breakers' timestamp extraction stops working after the Rule's configured **Future timestamp allowed** value (if set). All subsequent events received get the current time as their timestamp.

Workaround: Set a sufficiently large value for **Future timestamp allowed** – e.g., one year from the date you re/configure the rule. Alternately, restore normal timestamp extraction by restarting Worker Processes.

Fix: In Cribl Edge 4.1.1.

2023-02-05 – v.3.0.3–4.0.4 – `CRIBL_DIST_WORKER_PROXY` env var is ignored when Leader Node's Master URL is set via `instance.yml` [CRIBL-15540]

Problem: Setting the Cribl Edge Leader's Master URL via `instance.yml` causes Edge Nodes to ignore the `CRIBL_DIST_WORKER_PROXY` env var. Instead of trying to connect to the proxy configured in `CRIBL_DIST_WORKER_PROXY`, Edge Nodes will try to connect to other entities (e.g., Leader Nodes) directly, producing unexpected results.

Workaround: Set the Cribl Edge Leader's Master URL via the `CRIBL_DIST_MASTER_URL` env var, rather than via `instance.yml`.

Fix: In Cribl Edge 4.1.

2023-02-03 – v.4.0.0–4.0.4 – Metrics blocklist can't be changed on a global level [CRIBL-15081]

Problem: The Metrics blocklist cannot be modified from **Settings** > **Global Settings** > **General Settings** > **Limits**.

Workaround: Edit the `limits.yml` file's `metricsFieldsBlacklist` element. Add the event fields for which you want to disable metrics collection; remove any event fields for which you want to restore metrics collection.

Fix: In Cribl Edge 4.1.

2023-01-26 – v.3.5.0–4.0.3 – Preview Full > Send Out does not capture events [CRIBL-14914]

Problem: Selecting a Sample Data file's **Preview Full** > **Send out** option captures no events. This bug has been observed when capturing on Destinations.

Fix: In Cribl Edge 4.0.4.

2023-01-25 – v.3.5.x–4.3.1 – Cascading problems when Windows Event Logs Source collects from `ForwardedEvents` channel [CRIBL-14869]

Problem: Using the Windows Event Logs Source to collect events from a Windows `ForwardedEvents` channel can cause misattribution of logs to the local machine that hosts Cribl Edge, and undercollection of local logs from other logging channels (such as Security, System, or Application).

Workaround: Exclude the `ForwardedLogs` channel from your **Event Logs** selection.

Fix: In Cribl Edge 4.4.

2023-01-17 – v.4.0.x–4.0.4 – GitOps Push/read-only confirmation banner has dead link [CRIBL-14681]

Problem: After you enable GitOps `Push` mode, the resulting red confirmation banner contains a dead link labeled **GitOps Workflow**.

Workaround: To switch off `Push` mode, navigate to **Settings** > **Global** > **Git Settings**, and then set **GitOps workflow** back to `None`.

Fix: In Cribl Edge 4.1.

2023-01-12 – v.3.5.x through Current – Recently added Worker Nodes fail to appear on the Monitoring page [CRIBL-14627]

Problem: When Worker Nodes are added, there may be a delay before the Worker count is updated on the Monitoring page.

Workaround: Refresh your web browser.

Fix: Version TBD.

2023-01-11 – Multiple versions through 4.1.x – Kafka-based Sources' rebalancing is logged with exaggerated severity [CRIBL-14609]

Problem: The Kafka, Azure Event Hubs, and Confluent Cloud Sources log `REBALANCE_IN_PROGRESS` events at the `error` level, even though only **frequent** rebalancing indicates a system-level or processing issue.

Workaround: Treat infrequent rebalancing events as `warn`.

Fix: Depends on a change to the underlying kafkajs library.

2023-01-10 – v.3.5.x through Current – Configuration updates not shared between primary and standby Leaders [CRIBL-12814, CRIBL-14556]

Problem: With high availability/failover enabled, updates to the active Leader's configuration are not automatically synced to the standby Leader. Workers might not be able to connect to the standby Leader when it takes over.

Workaround: Use the filesystem to explicitly sync the updated `instance.yml` file across the two Leaders' hosts.

Fix: In – v.4.0.3, as an intermediate fix, enabling HA will prevent further UI-based changes to **Distributed Settings**. This will enforce config changes via edits to portable `instance.yml` files. Version TBD for automatic synchronization between Leaders' hosts.

2023-01-02 – v.4.0.0–4.0.2 – Google Cloud Chronicle Destination drops changes in distributed deployments with custom log types [CRIBL-14453]

Problem: A new Google Cloud Chronicle Destination with a custom log type might fail to save changes. This issue affects only distributed deployments.

Fix: In Cribl Edge 4.0.3.

2022-12-15 – v.3.5.4–4.0.3 – Changes to a Lookup table on the Leader don't always propagate to Edge Nodes [CRIBL-14299]

Problem: Modifying a Lookup table on the Leader doesn't always propagate the changes to Edge Nodes, even after clicking **Commit** and **Deploy**.

Workaround: Manually restart the Edge Nodes to refresh Lookup tables.

Fix: In Cribl Edge 4.0.4.

2022-12-13 – v.4.0.0 through Current – Default commit message missing for non-admin users [CRIBL-14239]

Problem: For users who have Roles as high as `owner_all`, but not `admin`, the Commit modal fails to display any **Default commit message** saved in **Git Settings**.

Workaround: Enter (or paste) a message per commit.

Fix: Version TBD.

2022-12-09 – v.4.0.1–4.0.2 – Users assigned the `owner_all` role cannot perform commits [CRIBL-14180]

Problem: When a user with the `owner_all` role tries to perform a commit, the commit fails with the UI displaying a `Forbidden` modal.

Workaround: Modify the `GroupFull` policy, as follows:

1. As a user with the `owner_all` role, try a `POST /version/commit` API call with a request body of `{ "message": "test: hello" }`. The commit should fail, and the request payload should be `{ message: "test: hello" }`.
2. If `$CRIBL_HOME/local/cribl/policies.yml` does not exist, copy `$CRIBL_HOME/default/cribl/policies.yml` to `$CRIBL_HOME/local/cribl/policies.yml`.
3. If your OS is Linux, run the command `chmod 0744 policies.yml`.
4. In `$CRIBL_HOME/local/cribl/policies.yml`, edit the `GroupFull` policy to match the following:

```
GroupFull:
  args:
    - groupName
  template:
    - PATCH /master/groups/${groupName}/deploy
    - GroupEdit ${groupName}
    - POST /version/commit
    - GET /version
    - GET /version/*
```

5. Now retry the `POST /version/commit` API call, again with a request body of `{ "message": "test: hello" }`. The commit should succeed, and the request payload should be `{ message: "test: hello", "group": "default", "effective": true }`.

Fix: In Cribl Edge 4.0.3.

2022-12-08 – v.4.0.1–4.0.2 – Landing page not displaying system metrics from Edge Nodes when teleporting from the Leader [CRIBL-14169]

Problem: When you teleport from a Leader running v.4.0.1 or 4.0.2 to an Edge Node that's running v.4.0.0 or earlier, the system metrics for the landing page will not populate. This also affects the system metrics shown in the Node drawer for the honeycomb and Node List View pages.

Workaround: Upgrade Edge Nodes to v.4.0.2 or higher.

Fix: In Cribl Edge 4.0.3.

2022-12-07 – v.4.0.1 – Google Cloud Pub/Sub Source and Destination can break on field validation [CRIBL-14160]

Problem: The Google Cloud Pub/Sub Source and Destination strictly validated entries in the **Topic ID** and **Subscription ID** fields. If you entered a full path (rather than the expected ID substring) in these fields, this strict validation broke the integration and broke the UI.

Workaround: Trim **Topic ID** and **Subscription ID** field values to just the ID.

Fix: Validation is rolled back in Cribl Edge 4.0.2.

2022-12-01 – v.4.0.2 – CrowdStrike FDR Source needs 6-hour visibility timeout [CRIBL-14067]

Problem: With its default **Visibility timeout seconds** setting of `600` (10 minutes), the CrowdStrike FDR Source can accumulate large backlogs when pulling data from Crowdstrike buckets.

Workaround: Set the **Visibility timeout seconds** to `21600` (6 hours), as CrowdStrike recommends. Leave both **Max messages** and **Num receivers** at their default `1` settings.

Fix: In Cribl Edge 4.0.3.

2022-12-01 – v.4.0.0 – Amazon S3 Source stops receiving data after upgrade [CRIBL-14093]

Problem: Upgrading to Cribl Edge 4.0.0 can cause the Amazon S3 Source to stop receiving data. This is due to a race condition between (premature) SQS messaging versus the Source's initialization.

Workaround: Skip v.4.0.0 (or temporarily roll back to v.3.5.4).

Fix: In Cribl Edge 4.0.1.

2022-11-30 – v.4.0.2 – QuickConnect-configured Sources' misleading "Enabled" status while disconnected [CRIBL-14041]

Problem: As you configure a new Source from the QuickConnect UI, the Source's **Enabled** slider will initially switch to `Yes`. This is misleading, because the Source is not yet connected to a Destination, so no data can flow.

Workaround: Save the Source in its config drawer. This resets the **Enabled** slider to its accurate `No` status, until you connect the Source to a Destination.

Fix: In Cribl Stream 4.0.3.

2022-11-28 v.4.0–4.0.2 – "Unknown config version" error can appear when deploying changes [CRIBL-13910]

Problem: Attempting to commit and deploy a change can trigger errors of the form `Unknown config version: "[hash]"`.

Workaround: On the Leader's host, upgrade the `git` client to v.1.9.1 or later.

Fix: In Cribl Edge 4.0.3.

2022-11-28 – v.4.1.0 through Current – Diag bundles might fail to download from teleported Edge Nodes [CRIBL-13999]

Problem: When you're remotely accessing a Edge Node's UI, diag bundles might fail to download from **Global Settings** > **Diagnostics**.

Workaround: Refresh the page and **Export** the bundle again. Alternatively, log directly into the Edge Node's UI before creating the diag.

Fix: Version TBD.

2022-11-23 – All versions through 4.1.0 – Edge Nodes show incorrect configurations after upgrade or commit/deploy from Leader [CRIBL-13863, CRIBL-15507]

Problem: When you commit and deploy changes from the Leader, Edge Nodes sometimes fail to automatically restart with the correct configuration changes.

Workaround: Manually restart the Edge Nodes.

Fix: In Cribl Edge 4.1.1.

2022-11-23 – v.4.0.0 – Teleported Edge Nodes don't display Distributed Settings or Diagnostics Settings [CRIBL-13868]

Problem: When displayed via remote access from the Leader ("teleporting"), a Worker's/Edge Node's **Settings** UI omits the **Distributed Settings** and **Diagnostics** left-nav links. This blocks remote access to features like configuring TLS communications between Edge Node and Leader.

Workaround: Access the Worker's UI directly on its host's port 9000 (or other configured port). As a precondition, you might need to undo any Disable UI Access setting on the parent Fleet.

Fix: In Cribl Edge 4.0.1.

2022-11-22 – v.3.2.0–4.0.0 – Chain Function creates extra Pipelines and Functions during initialization [CRIBL-13860]

Problem: Under certain circumstances, when initializing, the Chain Function created multiple unneeded Pipelines and Functions.

Fix: In Cribl Edge 4.0.1.

2022-11-21 – v.4.0.0 – Metrics service unavailable [CRIBL-13826]

Problem: The Monitoring and Fleet pages don't load, due to the systemd-tmpfiles service cleaning some of the Cribl socket files from the `/tmp/` directory.

Workaround: Stop the host system from cleaning the socket files from the `/tmp/cribl-*` directory. For example, on an Amazon EC2 instance, add a new file to `/etc/tmpfiles.d` with the line `X /tmp/cribl-*`. Then restart the tmpfiles cleanup service with: `systemctl restart systemd-tmpfiles-clean.service`. Finally, restart the Cribl server.

Fix: This issue no longer applies to Cribl.Cloud as of 4.0.1. On-prem customers should use the workaround, which will be added to the deployment documentation as a standard practice.

2022-11-18 – v.4.0.0–4.0.3 – License usage visible only for current day [CRIBL-13811]

Problem: In the listed versions, selecting **Monitoring** > **System** > **License** displays usage only for the current day.

Fix: In Cribl Edge v.4.0.4.

2022-11-16 – versions 3.4.0–4.0.4 – Source PQ in Smart mode can trigger excessive memory usage and OOM failures [CRIBL-13761]

Problem: Enabling Source-side Persistent Queues in Smart mode can trigger excessive memory usage, leading to out-of-memory failures and Worker Node instability.

Workaround: If you encounter OOM failures (most likely with high throughput), set PQ to `Always On` mode instead.

Fix: In Cribl Edge 4.1.

2022-11-14 – version 4.0.0 – `Forbidden` error banners mistakenly displayed to non-admin users [CRIBL-13681]

Problem: Non-admin users (with the `reader_all` Role) are mistakenly shown a continuous string of `Forbidden` error banners.

Workaround: Available upon request (documented in the ticket), but cumbersome. Cribl recommends upgrading instead.

Fix: In Cribl Edge 4.0.1.

2022-11-13 – v.4.0.0–4.0.2 – Edge/Kubernetes Logs Source repeats logs due to incorrect timestamp [CRIBL-13900]

Problem: The Kubernetes Logs Source is collecting redundant information for containers that don't emit their own timestamps. This Source assumes "current time" when the timestamps are missing, causing it to incorrectly stream older logs during restarts.

Fix: In Cribl Edge 4.0.3.

2022-11-08 – v.4.0.0 – Cribl Edge honeycomb display doesn't render null values [CRIBL-13595]

Problem: When you view Edge Nodes in **Map View**, a honeycomb displays values for each of the metrics you select in the **Measure** drop-down. When a value is null for a particular the Edge Node, the honeycomb should display `null`; instead, the value incorrectly renders as `ul i`.

Fix: In Cribl Edge 4.0.1.

2022-11-08 – v.4.0.0 – Broken in-app doc links [DOC-100]

Problem: The following Sources have broken in-app help links: Kubernetes Logs, Kubernetes Metrics, Windows Event Logs, and Windows Metrics.

Workaround: Use the above links to access these Sources' online documentation.

Fix: In Cribl Edge 4.0.1.

2022-11-07 – v.4.0.0 – Spurious “Secret decrypt failed with error" log messages appear after upgrade [CRIBL-13554]

Problem: After upgrading to Cribl Edge 4.0.0, multiple `Secret decrypt failed with error` messages might appear in logs. These are spurious, and you can disregard them.

Fix: In Cribl Edge 4.0.1.

2022-11-07 – v.4.0.0 – Upgrade status shown on Group Upgrade page doesn’t match Global Settings [CRIBL-13566]

Problem: The **Enable automatic upgrades** status shown on the **Settings** > **Group Upgrade** page doesn’t match the slider selection on the **Settings** > **Global Settings** > **Upgrade** tab.

Workaround: View the **Enable automatic upgrades** slider on the **Settings** > **Global Settings** > **Upgrade** tab to verify the current upgrade status.

Fix: In Cribl Edge 4.0.1.

2022-11-04 – v.3.5.1–4.0.4 – Avg Thruput (Bytes Per Second) not displayed for internal logs or metrics [CRIBL-13530]

Problem: No **Avg Thruput (Bytes Per Second)** data is displayed on Cribl Internal logs and metrics Sources’ **Charts** tab, nor for these Sources on their attached Routes. This is by design, because Cribl Edge does not perform BPS calculations for internal logs or metrics. But the visual disparity can be confusing.

Fix: In Cribl Edge 4.1.

2022-11-03 – v.4.0.0 – Sorting by column on Manage > Edge Node page results in blank list [CRIBL-13505]

Problem: When you click a column header on the **Manage** > **Edge Node** page, the list goes blank and rows are not sorted.

Workaround: Click the heading a couple of times to refresh the list.

Fix: In Cribl Edge 4.0.1.

2022-11-01 – All versions through Current – Source PQ in Smart mode could cause premature backpressure [CRIBL-13414]

Problem: Source-side persistent queuing enabled in `Smart` mode might trigger backpressure before the configured maximum queue size is reached. For TCP Sources, this will send backpressure back to the sender. For Sources using the UDP protocol, it will cause events to be dropped.

Fix: None, works as designed.

2022-11-01 – v.4.0.0 through Current – Default AppScope Config file can't be restored when in use [CRIBL-13400]

Problem: When the default AppScope Config file is assigned to an AppScope Source filter, you can't restore it to its default settings. You'll see an error message that the AppScope Config file is currently in use.

Workaround: On the AppScope Source's **AppScope Filter Settings** tab, delete the default AppScope Config from the **Allowlist**. Then, edit the AppScope Config Knowledge object to restore its default settings. You can now add this AppScope Config, with its restored settings, back to the AppScope Source's Allowlist.

Fix: Version TBD.

2022–10-20 – v.4.0.0 through Current – Pipelines with Clone and other Functions can show inconsistent total processing times [CRIBL-13125]

Problem: For Pipelines containing both a Clone Function and an async Functions (like Redis), the Pipeline Diagnostics modal's **Pipeline Profile** > **Process Time** graph will sum up the duration of all Functions' processing times. This sum can exceed the Pipeline's total duration displayed in the **Summary**.

Fix: Version TBD.

2022-11-03 – v.4.0.0 – Git settings don't immediately populate Commit modals [CRIBL-13501]

Problem: After you save Git settings (such as a default commit message), the new settings do not immediately appear in Commit/Git Changes modals.

Workaround: A hard or soft refresh will close the modal, but when you reopen it, it will reflect your new settings.

Fix: In Cribl Edge 4.0.1.

2022-10-12 – All versions through 3.5.4 – Event preview can hide some fields [CRIBL-12914]

Problem: Preview panes can hide some fields at the bottom of each event.

Workaround: Resize your browser window to give the preview pane more depth.

Fix: In Cribl Edge 4.0.

2022-10-11 – All versions through Current – GitOps Push mode doesn't support ad hoc Collection jobs [CRIBL-12868]

Problem: If you've enabled the GitOps Push workflow, you will be unable to run ad hoc Collection jobs.

Workaround: There are two options. 1. Temporarily disable the Push workflow in your environment. 2. Create a scheduled Collection job (with a relaxed cron schedule) on your dev branch, and push it to production through your Push workflow.

Fix: Version TBD.

2022-10-04 – All versions through 3.5.4 – File Monitor Source fails with high-volume logs and file rotation [CRIBL-12762]

Problem: The File Monitor Source might stop reading logs from an open file, if the log rotation service renames that file.

Workaround: Restart the File Monitor Source.

Fix: In Cribl Edge 4.0.

2022-10-03 – All versions through 4.0.4 – New Help drawers open under pinned Help drawers [CRIBL-12737]

Problem: If you've pinned a Help drawer open, it's possible to open additional Help drawers behind the pinned drawer.

Workaround: Close the pinned drawer to see the newly opened drawer.

Fix: In Cribl Edge 4.1.

2022-09-27 – All versions through 4.1.3 – Azure Event Hubs Destination with PQ drops events when inbound data is interrupted [CRIBL-17661, replaces CRIBL-12649]

Problem: When Azure Event Hubs (and other Kafka-based Destinations) have persistent queues (PQs) configured, an interruption of inbound data flow (e.g., due to network issues) can cause the Destination to start dropping events.

Workaround: On the Source whose events are being interrupted, configure an Always On persistent queue. This buffering will cause the Destination to drop fewer events. Note that persistent queuing will not engage until Cribl Edge has exhausted all attempts to send the data to the Kafka receiver.

Fix: In Cribl Edge 4.2.

2022–09-26 – All versions through 4.0.0 – Manage Groups/Fleets pages count Packs in Pipeline totals [CRIBL-12602]

Problem: On the **Manage Groups** and **Manage Fleets** pages, the **Pipelines** column overstates the actual number of Pipelines, because it includes Packs in each Fleet's total count.

Workaround: For each Fleet, click the link in its **Pipelines** column to see the Fleet's actual list of Pipelines.

Fix: In Cribl Edge 4.0.1.

2022-09-19 – v.4.0.0 – Subfleet search doesn't return parent Fleets [CRIBL-12465]

Problem: On the **Manage Fleets** page's **Fleets** tab, searching for a Subfleet name returns only the Subfleet, without information about its parent Fleets.

Fix: In Cribl Edge 4.0.1.

2022-09-06 – v.3.3.0–3.5.4 – Load-balanced Destinations with PQ enabled: Buffer flush errors and memory leaks [CRIBL-12203]

Problem: Destination logs might show errors of the form: `Attempted to flush previously flushed buffer token`. No data has been lost, but this will lead to a memory leak until the Edge Node is restarted.

This is most likely to occur when there are more available downstream hosts than a **Max connections** limit configured on the Destination.

Workaround: To prevent the bug from occurring, set **Max connections** to the default `0` (enabling unlimited connections). This will ensure that all discovered hosts enter the connection pool. If you encounter the bug, restarting the affected Edge Nodes (or letting an OOM the killer do so) will clear the memory leak.

Fix: In Cribl Stream 4.0.

2022-08-29 – All versions through 3.5.4 – Sources reject connections on ACME certs with no Subject [CRIBL-12097]

Problem: With multiple Sources, if you configure TLS mutual authentication using an ACME certificate with an empty `Subject` field, Cribl Edge will reject connections – even though RFC 6125 allows for `Subject` to be empty. This affects: Splunk Sources, Cribl internal Sources, Syslog, TCP, TCP JSON, Metrics, HTTP/S, Amazon Firehose, Elasticsearch API, DataDog, Grafana, Loki, Prometheus, and Windows Event Forwarder.

Workaround: Enter any value in the certificate's `CN` field. Any entry will cause CCribl Edge to match on the certificate's SAN (`subjectAltName`) extension.

Fix: In Cribl Edge 4.0.

2022-08-20 – v.3.2.0–4.0.3 – Deleting or modifying default Mapping Rule reclassifies Cribl-managed Cribl.Cloud Workers as hybrid Workers [CRIBL-11983]

Problem: Cribl.Cloud's UI currently allows deleting or modifying the `default` Mapping Rule. By doing so, an admin can inadvertently reclassify Cribl-managed Cribl.Cloud Workers as hybrid Workers. These might not be supported by your Cribl.Cloud plan.

Workaround: If you have an Enterprise plan, create a hybrid Worker Group to manage the resulting hybrid Workers.

Fix: In Cribl Edge 4.0.4.

2022-08-19 – v.3.5.1–3.5.4 – Using GitOps environment variables crashes the Leader [CRIBL-11972]

Problem: Where deployments rely on `CRIBL_GIT_*` environment variables, the Leader might crash on startup.

Workaround Restart the Leader, disabling the environment variables if necessary.

Fix: In Cribl Edge 4.0.

2022-08-25 – Versions 3.4.1 through 4.4.4 – Splunk Load Balanced Destination degradation with acks enabled [CRIBL-12066]

Problem: On a Splunk Load Balanced Destination, enabling the **Minimize in-flight data loss** (acknowledgments) field can cause high CPU drain and backpressure.

Workaround: On the Splunk LB Destination's **Advanced Settings** tab, disable **Minimize in-flight data loss**.

Fix: In Cribl Edge 4.5.

2022-08-16 – All versions through 4.3.0 – Splunk HEC shows higher outbound data volume than other Splunk Destinations [CRIBL-11922]

Problem: Events sent to the Splunk HEC Destination will show higher outbound data volume than the same events sent to the Splunk Single Instance or Splunk Load Balanced Destinations, which use the S2S binary protocol.

Fix: Version TBD.

2022-08-16 – v.3.1.2–3.5.2 – Auto timestamping randomly begins using current time [CRIBL-11925]

Problem: After a random interval, the auto timestamping feature might insert the current time into the `_time` field instead of using the timestamp from the event.

Workarounds: Restart the Stream instances to temporarily remedy the problem, or enable an Event Breaker rule's Manual Timestamp Format option.

Fix: In Cribl Edge 3.5.3.

2022-08-10 – v.3.5.1-3.5.2 – Changing Notification target type crashes New Target modal [CRIBL-11848]

Problem: Changing the **Target Type** while configuring a new Notification target crashes the **New Target** modal.

Workaround: Set the **Target type** once only while the modal is open.

Fix: In Cribl Edge 3.5.3.

2022-08-08 – v.3.x through Current – `cidr-matcher` does not support IPv6 addresses [CRIBL-11767]

Problem: The version of `cidr-matcher` supported by Cribl only supports IPv4 CIDR matching.

Fix: Version TBD.

2022-08-03 – v.3.5.0–3.5.1 – Cribl HTTP and Cribl TCP Sources are not enabled on Cribl.Cloud [SAAS-1905]

Problem: Cribl.Cloud's **Network Settings** >: **Data Sources** tab lists **Ingest Address** ports for the new Cribl HTTP and Cribl TCP Sources. However, these ports are not yet enabled.

Workaround: If you need these ports enabled before the next maintenance release, contact Cribl Support or your Solutions Engineer.

Fix: In Cribl Edge 3.5.2.

2022-08-02 – v.3.3.0–3.5.1 – Managed Edge Nodes mistakenly display "Unregistered" button [CRIBL-11652]

Problem: Managed Edge instances erroneously display an **Unregistered** button. Managed Edge Nodes cannot be registered, because they pull their registration/licensing information from the Leader. So this button should not appear.

Workaround: Ignore the scary button. Your managed Edge Node has inherited its Leader's registration.

Fix: In Cribl Edge 3.5.2.

2022-08-02 – v.3.x through 3.5.1 – Cribl.Cloud displays unsupported Sharing settings [SAAS-1899]

Problem: On Cribl.Cloud, Cribl Edge's **General Settings** > **Upgrade & Share Settings** display a **Sharing and live help** toggle, even though you cannot turn off telemetry on Cribl- or customer-managed (hybrid) Cribl.Cloud Workers.

Workaround: Ignore this toggle. The UI controls enable you to slide the toggle to `No` and save the result, but this will have no effect.

Fix: In Cribl Edge 3.5.2.

2022-07-21 – v.3.5.0–3.5.2 – Job Inspector shows higher Bytes In than Monitoring dashboards [CRIBL-11482]

Problem: The Job Inspector sometimes displays a higher byte count than other Monitoring dashboards display for the same collection job. The Job Inspector overstates the throughput because it disregards any Event Breakers and Filter expressions applied between initial collection and Pipelines.

Workaround: Rely on the Monitoring metrics to judge accurate data flow through Pipelines to downstream services.

Fix: In Cribl Edge 3.5.3.

2022-07-21 – v.3.4.0–3.5.1 – Data loss with Source-side persistent queueing enabled [CRIBL-11478]

Problem: With Source-side PQ enabled, events got stuck and did not process through Pipelines. The root cause was a Rename Function that used wildcards to rename required internal fields (specifically, `__pqSliceId` and `__offset`).

Workaround: Where Rename Functions potentially rename internal fields (especially via **Rename expressions** that include wildcards), either disable PQ on Sources that feed the parent Pipeline, or test and narrow the expression to affect only the desired fields.

Fix: In Cribl Edge 3.5.2.

2022-07-21 – v3.5.1-3.5.4 – Fake Pack appears after upgrading to v.3.5.1. [CRIBL-11481]

Problem: After an upgrade to v.3.5.1., a new Pack appears on the **Manage Packs** page. This Pack has no **Display Name** and no contents.

Workaround: Delete the ghost Pack.

Fix: Not observed as of Cribl Edge 4.0.

2022-07-20 – v3.5.0–3.5.1 – Manage Edge Nodes page fails to lazy-load Workers when scrolled [CRIBL-11474]

Problem: Scrolling the **Manage Edge Nodes** page fails to load more than 50 Edge Nodes.

Fix: In Cribl Edge 3.5.2.

2022-07-20 – v3.5.0–4.0.4 – Cribl Edge/Windows: Upgrading ignores existing mode (etc.) settings [CRIBL-11467]

Problem: When rerunning Windows installers on a system that hosts a previous Cribl version as a managed Edge Node, the installer does not read the distributed mode or other settings. The new version might install as a Leader instance.

Workaround: Run the new version's installer using the same mode and other options you used when installing the preceding version.

Fix: In Cribl Edge 4.1.

2022-07-19 – v.3.5.0–3.5.1 – Cribl Edge/Windows: Syslog Source disallows UDP binding [CRIBL-11439]

Problem: On Cribl Edge/Windows, attempting to bind the Syslog Source to a UDP port might trigger an error of the form: `bind EADDRINUSE 0.0.0.0:<port number>`. The reason is that Edge currently does not fully support the `socket.bind` on Windows.

Fix: In Cribl Stream 3.5.2.

2022-07-13 – v.3.5.0–3.5.1 – Cribl Edge: Fleet > List View tab doesn't correctly filter Edge Nodes [CRIBL-11183]

Problem: A Fleet Home page's **List View** tab doesn't correctly filter Edge Nodes. This tab displays all the Edge Nodes across all Fleets, instead of only those assigned to the Fleet.

Workaround: The **Map View** page is filtered correctly.

Fix: In Cribl Edge 3.5.2.

2022-07-04 – All versions through v.3.5.3 – No extended characters in Publish Metrics Function > Event field name field [CRIBL-8968, CRIBL-10984]

Problem: A Publish Metrics Function's **Event field name** values should contain only letters, numbers, underscores (_), and `.` characters (to separate names in nested structures). Using other characters will cause the parent Pipeline to stop sending events. Cribl Edge 3.5.0 and above check for valid characters when you save the Function's configuration, but in prior versions, the invalid config will save without an error message – the failure will happen at runtime, with errors echoed only in the logs.

Workaround: Ensure that **Event field name** values contain no extended characters.

Fix: In Cribl Edge 3.5.4.

2022-07-01 – All versions through 3.5.0 – S3-based Sources' unread Region triggers auth errors [CRIBL-10981]

Problem: On Amazon S3–based Sources, if you concatenate the AWS Region into the **Queue** field's URL or ARN, the SQS client is correctly configured to that region, but the S3 client is not. You will see authentication errors of the form: `The AWS Access Key ID you provided does not exist in our records.`

Workaround: Explicitly set the **Region** drop-down, even if the URL/ARN already contains the same Region.

Fix: In Cribl Edge 3.5.1.

2022-06-30 – All versions through Current – Persistent queue with compression requires oversize max queue limit [CRIBL-10965]

Problem: If you configure persistent queueing to both enable **Compression** and set a **Max queue size** limit, set that limit optimistically and monitor the results. Due to an error in computing the compression factor, Cribl Edge will not fully use **Max queue size** values below or equal to the volume's available disk space. This can lead to a mostly empty disk, lost data (either blocked or dropped), and/or excessive backpressure.

Workaround: With **Compression** enabled, set the **Max queue size** to a value **higher** than the volume's total available physical disk space (disregarding compression). Alternatively, leave **Max queue size** empty, to impose no limit. Monitor overall throughput and the queue size (if PQ engages), to verify that Cribl Edge is maximizing the queue.

Fix: Version TBD.

2022-06-29 – v.3.4.2–3.5.1 – Can't install a new Cribl Edge instance as a named user [CRIBL-10907, CRIBL-11457]

Problem: Bootstrapping a new Leader or Worker with a command of this form fails, with no systemd service created: `cribl boot-start enable -u <username>`

The symptom is an error of the form:
`error: found user=0 as owner for path=/opt/cribl/local, expected uid=1001.` This does not affect upgrades to existing instances.

Workaround: Issue the `[sudo] chown -R <username>: /opt/cribl` command a second time. Then reattempt the `cribl boot-start` command.

Fix: In Cribl Edge 3.5.2.

2022-06-28 – v.3.5.x through Current – Cribl Edge/Windows Limitation: Cannot upgrade Edge Nodes from the Leader's UI [CRIBL-10870]

Problem: Cribl Edge/Windows does not support upgrading Edge Nodes via the Leader's UI.

Workaround: Rerun the Windows installer on each Edge Node, specifying the same installation options/parameters you used when installing the preceding version.

Fix: In Cribl Edge 4.3.

2022-06-27 – v.3.3.0–3.5.1 – Edge vs. Stream conflict in `/tmp` directory [CRIBL-10806]

Problem: If you install Cribl Edge and Cribl Stream on the same machine or VM – e.g., running Edge as `root` and Stream as a `cribl` user – both services attempt to use the `tmp/cribl_temp` subdirectory. Starting Stream after Edge will throw an error of the form: `EPERM: operation not permitted, rmdir '/tmp/cribl_temp'`. Starting Edge after Stream will change the folder's ownership, blocking subsequent restarts of Stream.

Workaround: For either Cribl Edge or Cribl Stream, set the `CRIBL_TMP_DIR` variable to a separate base subdirectory like `/tmp/stream/` or `/tmp/edge/` before starting the app.

Fix: Cribl Edge 3.5.2 and later set separate `tmp/` subdirectories by default.

2022-06-24 – v.3.5.0 – Leader takes excessive time to report healthy status [CRIBL-10768, CRIBL-11127]

Problem: Upgrading from v.3.4.x to v.3.5.0 tripled the latency before some Leaders reported healthy status. The presumed cause was that load time increased due to an accumulation of persistent metrics, which progressively increased with more Worker Nodes and more uptime.

Workaround: If upgrading is not possible or insufficient, move the existing metrics subdirectory away from `$CRIBL_HOME/state/metrics/`. Cribl Edge will re-create that as an empty subdirectory, and begin accruing fresh metrics there.

Fix: In Cribl Stream 3.5.1.

2022-06-23 – v.3.5.0 – Logs not viewable at Worker Group level [CRIBL-10665]

Problem: If you select **Monitoring** > **Logs**, and then select a Worker Group from the drop-down at upper left, you might see no logs at the Group level. Instead, you will see an error banner of the form: `ENOENT: no such file or directory, scandir '/opt/cribl_data/failover/log/group/<group-name>'`

Workaround: From the drop-down at upper left, select individual Worker Nodes to view their logs.

Fix: In Cribl Edge 3.5.1.

2022-06-09 – v.4.0.x–4.1.0 – Edge Settings mistakenly display Process Count controls [CRIBL-10402]

Problem: Edge Fleets' **Fleet Settings** > **Worker Processes** page incorrectly includes editable **Process Count** and **Minimum Process Count** controls. Each Edge Node is locked to `1` Worker Process, so changes to these fields' values will have no effect.

Workaround: Ignore these two fields.

Fix: In Cribl Edge 4.1.1.

2022-06-01 – v.3.4.2-3.5.0 – Bootstrapping fails to create SystemD service file, or starts wrong service [CRIBL-10250]

Problem: Bootstrapping a 3.4.2 Worker fails. The terminal displays no `Enabled Cribl to be managed by ...` confirmation message.

Workaround: Explicitly run the `boot-start` CLI command.

Fix: In Cribl Edge 3.5.1.

2022-05-31 – v.3.4.2 – Cloning/loading Groups page breaks with `Config Helper service is not available...` error [CRIBL-10228, CRIBL-10229]

Problem: After cloning a Group containing at least one Pack, the UI hangs with a spinning pinwheel and an error banner reading `Config Helper service is not available...`. The root cause is that cloning the Group failed to copy over the Pack, leaving the new Group with broken references to it.

Workaround: Use the filesystem to manually copy missing Packs from the original Group to the cloned Group. If these Packs' contents are not needed in the new Group, it's sufficient to just create the parallel subdirectory with: `mkdir $CRIBL_HOME/groups/$destGroup/default/$packName`.

Fix: In Cribl Edge 3.5.

2022-05-19 – v.3.3.0–3.5.1 – System Metrics Source skips filesystem events on LVM volumes [CRIBL-10092]

Problem: When monitoring LVM logical volumes, the System Metrics Source omits `node_filesystem_*` events.

Fix: In Cribl Edge 3.5.2.

2022-05-06 – v.3.5.0-3.5.4 – Running ~1,000 Edge Nodes crashes Leader [CRIBL-9859]

Problem: Attempting to bring up approximately 1,000 Edge Nodes (even in small batches) can crash the Leader.

Workaround: To support up to 1,000 Nodes, apply this Node.js setting:
`export NODE_OPTIONS=--max-old-space-size=8192`

Fix: In Cribl Edge 4.0.

2022-04-28 – v.3.4.0–3.4.1 – REST Collector Misinterprets `.` character in Response Attributes [CRIBL-9708]

Problem: Where a REST Collector's Response Attributes contain a `.` character, it misinterprets this as an attribute nested within an object, not a literal character.

Workaround: If possible, use other Response Attributes to fetch the next page. Otherwise, skip this version, or downgrade to a known well-behaved version.

Fix: In Cribl Edge 3.4.2.

2022-04-27 – v.3.4.1 – Git Changes drop-down returns `Invalid Date` [CRIBL-9698]

Problem: After opening the left nav's **Changes** fly-out, the **Version** drop-down can display `Invalid Date` instead of commits' dates.

Workaround: Upgrade your installed `git` client to v.2.1.1 or newer; or skip this Cribl Edge version.

Fix: In Cribl Edge 3.4.2.

2022-04-26 – v.3.4.1 – HTTP 500 error when git is absent [CRIBL-9684]

Problem: A `500–Internal Server Error` error banner can indicate that `git` is not installed on Cribl Edge's host.

Workaround: Install `git` on the Leader or single instance. On affected Edge Nodes, install `git`, followed by a `git init` and an empty push to a new `master` branch. A simpler workaround for Edge Nodes is to just enable remote access (Stream, Edge) from the Leader.

Fix: In Cribl Edge 3.4.2.

2022-04-22 – v.3.3.0-3.4.1 – Datadog Agent API key ignored when forwarding metrics to Datadog [CRIBL-9633]

Problem: On a Datadog Destination, when **Allow API key from events** is set to `Yes`, the API key from a Datadog Agent Source is not correctly emitted with metric events sent to Datadog. Instead, Cribl Edge simply sends the static **API key** configured in the Destination. (This affects metrics events only; other data types correctly forward the event's API key.)

Workaround: Set **Allow API key from events**. to `No`. Configure the desired API key in the Destination's static **API key** field. (If you need multiple API keys, can create multiple Destinations, and route events to each based on the `__agent_api_key` field value. This value contains the Agent API key per event.)

Fix: In Cribl Edge 3.4.2.

2022-04-21 – v.3.4.1 – Failed restart, `Uncaught (in promise)` error, after changing Authentication settings [CRIBL-9614]

Problem: After changing Authentication settings, or configuring an external auth provider, Cribl Edge might fail to restart. Indications are errors of the form: `Uncaught (in promise) TypeError: Cannot read properties of undefined (reading 'workerRemoteAccess')`, or: `Uncaught (in promise) TypeError: l.api is undefined`.

Workaround: Directly edit the `cribl.yml` file's `auth:` section.

Fix: In Cribl Edge 3.4.2.

2022-04-21 – All versions – Cribl.Cloud login page distorted on iPad [SAAS-1141]

Problem: On certain iPads, we've seen the Cribl.Cloud login page's left text column repeated twice more across the display. These unintended overlaps prevent you from selecting (or tabbing to) the **Log in with Google** button.

Workaround: If you encounter this, the only current workarounds are to either use Google SSO on a desktop browser, or else use a different login method.

Fix: No planned fix.

2022-04-19 – All versions through Current – Upgrade Version List is Limited to Latest Version [CRIBL-9568]

Problem: The **Choose Version** drop-down list displays the latest version only. This creates an issue when you want to incrementally upgrade between prior versions.

Fix: Version TBD.

2022-04-19 – v.3.3.0–3.4.2 – Spurious buffer token flush error in TCP-based Destinations with PQ enabled [CRIBL-9565]

Problem: This error message can mistakenly appear in logs: `Attempted to flush previously flushed buffer token`. You can ignore it on TCP-based, load-balanced Destinations (Splunk Load Balanced, TCP JSON, Syslog/TCP, and Cribl Stream) with persistent queueing enabled.

Fix: In Cribl Edge 3.5.

2022-04-18 – v.3.3.1–4.0.0 – Lookup Functions intermittently fail [CRIBL-9539]

Problem: Lookup Functions within some Pipelines were skipped up to ~20% of the time. Restarting Cribl Edge resolves this temporarily, but the failure eventually resurfaces as the new session proceeds.

Workaround: Where a Lookup Function fails, substitute an Eval Function, building a ternary JS expression around a `C.Lookup` method.

Fix: In Cribl Edge 4.0.1.

2022-04-15 – All versions through 3.4.1 – Git permission errors [CRIBL-9530]

Problem: Multiple Linux distro's have backported a permissions-restriction patch from `git-2.35.1` to earlier versions, notably including Ubuntu 20.04 LTS. If you see errors of the form `fatal: unsafe repository ('/opt/cribl' is owned by someone else)` on the command line or in the Cribl Edge UI, this indicates an ownership mismatch (current user versus file base) on the directory corresponding to `$CRIBL_HOME` or `$CRIBL_VOLUME_DIR`.

Workaround: Use `chown` to recursively set permissions on all files in `/opt/cribl/` (or in or `$CRIBL_VOLUME_DIR`'s target directory) to match the user running Cribl Edge.

Fix: In Cribl Edge 3.4.2.

2022-04-15 – v.3.4.0 – Workers report incomplete metrics [CRIBL-9524]

Problem: After upgrading to v.3.4.0, Worker Nodes failed to report all metrics. Missing metrics were logged with a warning of the form: `failed to report metrics`, and with a reason of the form: `Cannot read property 'size' of undefined`.

Fix: In Cribl Edge 3.4.1.

2022-04-14 – v.3.4.0-3.4.1 – Monitoring visualizations lack color indicators [CRIBL-9510]

Problem: When you hover over the Monitoring page's graphs, the expected red/yellow/green status indicators are missing.

Fix: In Cribl Edge 3.4.2.

2022-04-07 – v.3.4.0–3.4.1 – Chain Function referencing a Pack triggers errors on Pack's Functions [CRIBL-9235]

Problem: After you define a Chain Function that references a Pack, Functions on Pipelines within that Pack will throw errors of the form: `Failed to load. Function <Function-name> is missing. Please fix, disable, or remove the Function.` You might also find that you cannot add more Functions within the Pack. (However, data might continue to flow, despite these errors.)

Workaround: Remove the Chain Function, or its Pack reference. Refactor your flow logic to avoid this combination.

Fix: In Cribl Edge 3.4.2.

2022-04-05 – v.3.4.0 through Current – Upgrade to 3.4.0 disables Worker/Node UI access [CRIBL-9175]

Problem: Upon upgrading to v.3.4.0, even if the (prior, global) Worker/Node **UI access** option was set to `Yes`, the new per–Fleet toggles are all set to `No` by default.

Workaround: On the **Manage Fleets** page, re-enable the **UI access** toggles as desired. If these toggles are not displayed (with a Free license), edit `groups.yml` to add the key-value pair `workerRemoteAccess: true` within each desired Fleet.

Fix: Version TBD.

2022-03-29 – v.3.4.0 – `this.dLogger.isSilly is not a function` errors [CRIBL-9019]

Problem: Errors of this form have been observed with multiple triggers: 1. Disabling a customized Source or Destination (it remains enabled). 2. Importing and deleting a Pack. 3. Changing a channel's logging level.

Workaround: Delete and re-create affected Sources and Destinations. (No workaround has been identified for the Pack or logging errors.)

Fix: In 3.4.1.

2022-03-23 – v.3.4.0 – Monitoring page error in distributed environments with many Edge Nodes [CRIBL-8938]

Problem: The Monitoring page displays an error where distributed environments have more than 30 Edge Nodes.

Fix: In 3.4.1.

2022-03-23 – Collect parameters values not evaluated as expressions [CRIBL-8932]

Problem: In a REST Collector's **Collect parameters** table, entering a parameter name like "earliest" as plain text will cause it to be interpreted as a literal string. This will not be evaluated as the `earliest` time-range parameter.

Workaround: Backtick the parameter name, e.g.: `` `earliest` ``.

Fix: Tooltip clarified in Cribl Edge 4.0.

2022-03-23 – Upgrading v.3.3.1 Leader to v.3.4.0 blocks upgrade to Workers [CRIBL-8918]

Problem: When you explicitly upgrade a Leader Node to 3.4.0, you can't push upgrades to Workers.

Workaround: Automatic upgrades work as expected. At **Settings** > **Global Settings** > **System** > **Upgrade**, set **Disable Automatic upgrades** to `No`.

Fix: In 3.4.1.

2022-03-17 – All versions through 3.4.0 – Failed systemd restarts due to erroneous `ExecStopPost` command [CRIBL-8807]

Problem: On systemd, the Cribl service can stop with an unsuccessful return code. This was caused by a malformed `ExecStopPost` command in Cribl's provided `cribl.service` file.

Workaround: In `cribl.service`, delete the whole `ExecStopPost=...` line, and change the `Restart` parameter's value from `on-failure` to `always`. (The first deletion necessitates the second change.)

Fix: Versions 3.4.1 and later install a `cribl.service` file incorporating both the above changes. **If you are upgrading from v.3.4.0 or earlier, and you notice dead Workers, check and update your existing `cribl.service` configuration to match the changes above.**

2022-03-17 – v.3.4.0 – Source PQ re-engages while draining, causing blockage/backpressure [CRIBL-8800]

Problem: A Source with Persistent Queueing enabled can mistakenly re-engage queueing while still draining its queue – leading to blocked throughput and backpressure. This occurs after a Destination goes offline and back online, once or more, leaving the Source in an undetermined queue state.

Workaround: Wait 60 seconds for another queue drain event, to see if PQ behavior goes back to normal. If the problem persists, restart Cribl Stream.

Fix: In Cribl Edge 3.4.1.

2022-03-17 – v.3.4.0 – Monitoring pages break without git [CRIBL-8799]

Problem: Unless git is installed locally, two Monitoring pages fail to display, with errors of the form `Versioning not supported`. The pages are **Monitoring** > **Overview** and **Monitoring** > **System** > **Licensing**. To resolve the errors, git must be installed even on single-instance deployments.

Workaround: Install git.

Fix: In 3.4.1.

2022-03-16 – v.3.2.2-3.4.0 – Upgrading from earlier versions degrades performance [CRIBL-8784]

Problem: After upgrading to any of the indicated versions, customers with active Splunk Destinations noticed that CPU load increased, triggering backpressure more readily.

Workaround: Skip these versions, or downgrade to a known well-behaved version.

Fix: In 3.4.1.

2022-03-11 – v.3.4.0 – In Free versions, Worker/Node UI access can't be accessed via UI [CRIBL-8707, CRIBL-8945]

Problem: In v.3.4, Cribl moved the Worker/Node **UI access** toggle from global **Settings** > **System** > **Distributed Settings** to per–Fleet toggles on the **Manage Fleets** page. But with a Free license, these controls aren't displayed in the UI at all.

Workaround: Directly edit `groups.yml` to add the key-value pair `workerRemoteAccess: true` within each desired Fleet.

Fix: In 3.4.1.

2022-03-08 – v.3.3.1 through Current – GitOps + License expiration = Catch-22 [CRIBL-8600]

Problem: If your Enterprise license expires while you have enabled the GitOps Push workflow, you will encounter the following block: Cribl Stream is in read-only mode, triggering a `Forbidden` error when you try to update your license key. But you also cannot reset the workflow from `Push` to `None`, because the expired license disables GitOps features.

Workaround: Contact Cribl Support for help updating your license.

Fix: Version TBD.

2022-03-07 – v.3.2.x-3.4.x – Undercounted `total.in_bytes` dimension metrics [CRIBL-8572]

Problem: The `cribl.logstream.total.in_bytes` dimension sent to Destinations can show a lower data volume (against license quota) than `cribl.logstream.index.in_bytes`. This latter `index.in_bytes` dimension displays the correct license usage, and matches what's reported on the **Monitoring** dashboard.

Workaround: Within the `cribl_metrics_rollup` Pipeline that ships with Cribl Stream, disable the Rollup Metrics Function.

Fix: Couldn't reproduce the error.

2022-03-03 – v.3.3.x – Elasticsearch API Source (distributed mode) stops receiving data after upgrade to 3.3.x [CRIBL-8515]

Problem: After upgrading a distributed deployment to version 3.3.x, the ElasticSearch API Source might stop receiving data. This occurs when your existing Source config's **Authentication type** was set to `Auth Token`. The root cause is a 3.3.x schema change that unset this auth type to `None`.

Workaround: In the Elasticsearch API Source configuration, reset the **Authentication type** to `Auth Token`. Then commit and deploy the restored config.

Fix: In Cribl Stream 3.4.

2022-02-24 – v.3.3.x – After upgrade to 3.3.0, ElasticSearch Destination can't write to indexes whose name includes – [CRIBL-8451]

Problem: After upgrade to version 3.3.x, LogStream cannot send data to Elasticsearch indexes whose name includes a hyphen (-).

Workaround: In the Elasticsearch Destinations's **Index or Data Stream** field, surround these index names in "quotes" or backticks.

Fix: LogStream 3.3.1 added format validation and error-checking.

2022-02-18 – v.3.3.x – Worker/Leader communications blocked by untrusted TLS certificate after upgrade to 3.3.0

[CRIBL-8361]

Problem: With TLS enabled on Worker/Leader communications, after an upgrade from LogStream 3.2.x to v.3.3.x, Workers might fail to communicate with the Leader due to an untrusted SSL certificate on the Leader. Logs will show `connection errors` of the form: `"unable to verify the first certificate"` or `self signed certificate in certificate chain`.

Workaround: On each affected Worker: In the `instance.yml` file's' `distributed: master: tls:` section, set: `"rejectUnauthorized: false"`. Then restart Cribl Stream.

Fix: In 3.4.1.

2022-02-17 – v.3.3.0 – Do not configure InfluxDB Destination on LogStream 3.3.0 [CRIBL-8354]

Problem: Configuring InfluxDB Destinations can fail on LogStream 3.3.0. This problem has been observed only on distributed deployments, and data does flow to InfluxDB. But the InfluxDB config will not appear in LogStream's UI, and the broken config will block editing of other Destinations.

Workaround: If you need to send data to InfluxDB, skip v.3.3.0, and wait for the next release. If you encounter the broken UI, edit `outputs.yml` to remove any `influxdb_output` sections, and then restart LogStream. This will unblock UI-based editing of other Destinations.

Fix: In LogStream 3.3.1.

2022-02-17 – v.3.2.2-3.3.x – REST Collector pagination fails on duplicate/nested attribute names [CRIBL-8352]

Problem: The REST Collector does not consistently differentiate between multiple (nested) attributes that share the same name. This can break pagination that relies on a `Response Body Attribute`.

Workaround: If possible, select a different Pagination method, or specify a unique attribute name.

Fix: In Cribl Stream 3.4.

2022-02-17 – v.3.3.x – Can't upgrade Workers to 3.3.x via UI [CRIBL-8346]

Problem: Attempting to upgrade on-prem Workers to version 3.3.x via the UI can fail, with an error of the form: `Group <group-name> is a managed group`.

Workaround: Edit the Leader's `$CRIBL_HOME/local/cribl/groups.yml` file to delete all instances of the key-value pair `onPrem: false`. Then, resave the file and reload the Leader.

Fix: In Cribl Stream 3.4.

2022-02-10 – All versions, 3.2.1+ – Syslog Source's previewed data doesn't proceed through Routes [CRIBL-8210]

Problem: The Syslog Source's **General Settings** > **Input ID** field's default filter expression is `__inputId.startsWith('syslog:in_syslog:')`. The same filter appears in the **Preview Full** tab's **Entry Point** drop-down, except without the final colon. This means that data sent using **Preview Full**'s version of the filter will not go through Routes that use the **Input ID** version.

Workaround: When sending data from **Preview Full** to a Route, if both use the `syslog:in_syslog` filter, edit the Route's filter to remove the final colon. This will make the filter identical in both places, routing data correctly.

2022-02-10 – v.3.3.0-3.5.x – Slow-Mo Notification Notifications [CRIBL-8205]

Problem: After you create a Notification on a Destination, the new Notification might take up to several minutes to appear on the Destination config modal's **Notifications** tab.

Details: This delay has not been reproduced consistently. Some Notifications appear immediately.

Fix: In Cribl Edge 4.0.

2022-02-08 – v.3.3.x – Missing event from Datadog Agent v.7.33.0+ [CRIBL-8132]

Problem: If ingesting metrics from Datadog Agent 7.33.0 or later (i.e., you have set the `DD_DD_URL` environment variable or the `dd_url` YAML config key), LogStream's Datadog Agent Source will not ingest `agent_metadata` events. This is due to a breaking change in Datadog Agent 7.33.0, which split out part of the prior `/intake/` endpoint into a new `/api/v1/metadata` endpoint.

Workaround: Use Datadog Agent v.7.32.4 or earlier.

Fix: In Cribl Stream 3.4.

2022-02-08 – All versions through 3.3.x – Syslog and Metrics Sources' default filter expression needs correction [CRIBL-8115]

Problem: In the Syslog and Metrics Sources's **Logs** tab, the auto-generated filter expression (`channel =='input:in_syslog'`) is not specific enough and yields no results, because it doesn't address these Sources' two channels (`input:in_syslog:tcp` and `input:in_syslog:udp`).

Workaround: Replace this default filter expression with `channel.startsWith('input:<input_ID>')`. For example: `channel.startsWith('input:in_syslog:')`.

Fix: In Cribl Stream 3.4.

2022-02-07 – v.3.2.2-3.4.0 – Okta integration fails after upgrading from v.3.1.3 to v.3.2.2 [CRIBL-8105]

Problem: Okta (OpenID Connect) authentication on Cribl Stream fails behind a proxy, when using environment variables (`http_proxy` or `https_proxy`).

Workaround: Configure the proxy to work in transparent mode, to bypass the `http*_proxy` variables. If the proxy is required, leave the `User Info URL` empty, and Cribl Stream can obtain user details from the `JWT token`.

Fix: In 3.4.1.

2022-02-04 – v.3.2.2-3.4.0 – Mask Function slows down post-processing Pipeline [CRIBL-8043]

Problem: A post-processing Pipeline with a Mask Function can slow down drastically, showing high CPU usage on Workers.

Workaround: If practical, promote the same Mask Function to a pre-processing or processing Pipeline.

Fix: In Cribl Stream 3.4.1.

2022-2-03 – All versions though 3.4.0 – JSON Serialize Function in post-processing Pipeline won't process [CRIBL-8013]

Problem: A JSON Serialize Function in a post-processing Pipeline will throw errors of the form: `Failed to process event in Function: Maximum call stack size exceeded`.

Workaround: Promote the Serialize Function to a processing Pipeline, upstream from the Destination.

Fix: In 3.4.1.

2022-02-02 – v.3.0.0 through Current – Upgrading a Pack overwrites Lookup and sample-data customizations [CRIBL-7998]

Problem: Upgrading a Pack overwrites any customizations you've made to the Pack's original/default Lookup tables and sample-data files.

Workaround: Duplicate the Pack's default Lookup tables and sample-data files, and customize your duplicate copies. Upgrading the Pack will not overwrite your local copies.

Fix: Version TBD.

2022-01-18 – v.3.2.0-3.4.0 – Diagnostics > System Info page omits some entries [CRIBL-7731]

Problem: In current Cribl Stream versions, the **Diagnostics** > **System Info** page/tab omits certain statistics (`sysctl`, `ulimits`, `network stats`, etc.) that were previously displayed below the `cpus` and `nets` elements.

Workaround: The hidden information is nevertheless available within diagnostic bundles.

Fix: In Cribl Stream 3.4.1.

2022-01-18 – v.3.2.2 – Usernames containing certain letters cause some API requests to fail [CRIBL-7728]

Problem: Usernames containing certain letters can cause some API requests to fail with an `Invalid character in header content ["cribl-user"]` error. This can happen even when the username is valid for an Identity Provider.

Workaround: In your usernames, make sure that the letters you use are limited to the letters in Latin alphabet no. 1 as defined in the ISO/IEC 8859-1 standard.

Fix: In LogStream 3.3.

2022-01-11 – v.3.2.2 – Git Collapse Actions disaggregates [CRIBL-7638]

Problem: With **Git Settings** > **Collapsed actions** enabled, the combined **Commit & Push** button appears in the modal as expected, but then saving a new change splits it back into two separate buttons: **Commit** and **Deploy**.

Workaround: Disable the **Collapse actions** setting, then commit and deploy separately.

Fix: In LogStream 3.3.

2022-01-04 – v.3.2.x – Enabling GitOps deletes LogStream's `bin`, `logs`, and `pid`subdirectories [CRIBL-7513]

Problem: Enabling GitOps with LogStream 3.2.x, via either CLI or UI, can cause the deletion of the `bin`, `logs`, and `pid`subdirectories. This disables LogStream. The root cause is that `git` versions 2.13.1 and lower did not fully respect paths listed in `.gitignore`.

Workaround: Upgrade your `git` client to v.2.13.2 or higher, to correct the underlying behavior.

Fix: In LogStream 3.3.

2021-12-21 – v.3.2.2-4.0.x – Chain Function degrades CPU load and throughput [CRIBL-7445]

Problem: Chaining Pipelines via the Chain Function can increase CPU load, and can signficantly slow down data throughput.

Workaround: Consolidate all Functions – per processing scenario – into a single Pipeline.

Fix: In Cribl Edge 4.1.

2021-12-15 – v.3.2.1-3.2.2 – Backpressure when writing to S3 and other file-based Destinations [CRIBL-7364]

Problem: On file-based Destinations, enabling the **Remove staging dirs** toggle can trigger a race condition when inbound events per second reach a high rate. This leads to backpressure.

Workaround: Disable LogStream's native **Remove staging dirs** option. Instead, as the user running LogStream, set up a cron job like this:

```
crontab -e
0 1 * * * find <stagingdir> -type d -empty -mtime +1 -delete
```

Fix: In LogStream 3.3.

2021-12-13 – v.2.4.4-3.1.1 – Upgrades via UI delete sample files [CRIBL-7347]

Problem: Using the UI to upgrade LogStream versions prior to 3.1.2 will silently delete sample data files created by users. This affects using a Leader's UI (any version) to upgrade Workers running LogStream version 3.1.1 or earlier. It also affects using the UI to upgrade the Leader itself, or a Single-Instance deployment, from v.3.1.1 or earlier to any newer version.

Workarounds: 1. Upgrade pre-3.1.2 versions via the filesystem. 2. If you choose to upgrade pre-3.1.2 versions via the UI, first save to the filesystem any custom sample files that you want to preserve. (Or, to copy directly from the filesystem, LogStream stores sample files internally at `$CRIBL_HOME/data/samples`, and stores an index of all sample files in `/$CRIBL_HOME/local/cribl/samples.yml`.)

Fix: Does not affect Workers running Cribl Stream (LogStream) 3.1.2 or higher. Does not affect self-upgrades of Leaders or Single Instances running v.3.1.2 or higher.

2021-12-10 – All Versions – API Reference is temporarily unstyled

Problem: Our documentation's API Reference has temporarily lost some visual styling, while we swap in a new rendering component.

Workaround: Manually expand accordions, as needed.

Fix: Published as of 12/20/2021.

2021-12-09 – v.3.2.1 – File-based Destinations display an unavailable Persistent Queue option [CRIBL-7293]

Problem: File-based Destinations' **Backpressure behavior** drop-down misleadingly displays a `Persistent Queue` option, which is not really available on these Destinations. (Applies to Filesystem and some Amazon, Azure, and Google Cloud Destinations.) Selecting this option displays an error, and the configuration cannot be saved.

Workaround: Don't fall for clicking that fake `Persistent Queue` option.

Fix: In LogStream 3.2.2.

2021-12-07 – v.3.2.1 – System Metrics Source's documentation doesn't open in Help drawer

Problem: After clicking the System Metrics Source's Help link, the drawer displays the error: "Unable to load docs. Please check LogStream's online documentation instead." This Source's documentation is also missing from the 3.2.1 docs PDF.

Workaround: Go to the live System Metrics docs page.

Fix: In LogStream 3.2.1, as duplicate of CRIBL-6319.

2021-12-07 – v.3.1.2-3.1.3 – Increasing CPU load over time [CRIBL-7268]

Problem: CPU load increases over time, with a slow increase in memory usage. Restarting LogStream temporarily resolves these symptoms. The root cause is needlessly collecting a high volume of full-fidelity metrics from the CriblMetrics source.

Workaround: Disable the CriblMetrics Source.

Fix: Upgrade to LogStream 3.2.1, which provides an option to disable the CriblMetrics Source's **Full Fidelity** toggle.

2021-11-24 – v.3.2.0 through Current – With processing Pipelines in QuickConnect, Preview Full doesn't show Functions' results [CRIBL-7113]

Problem: If a processing Pipeline has been inserted between a QuickConnect Source and Destination, selecting the **Pipelines** page, and then selecting **Preview Full** with a data sample, doesn't show the Pipeline's effects on the **OUT** tab. This affects only Pipelines inserted in a QuickConnect connection line. (Attaching a pre-processing Pipeline to a QuickConnect Source doesn't inhibit Preview.)

Workarounds: 1. Remove the Pipeline from QuickConnect, and re-create the same Source -> Pipeline -> Destination architecture under **Routing** > **Data Routes**. 2. Rely on **Preview Simple**.

Fix: Version TBD.

2021-11-24 – v3.2.0 – Data from Collectors and Collector-based Sources isn't reaching Routes [CRIBL-7109]

Problem: Routes are not recognizing data from Collectors and from the following Collector-based Sources: Prometheus Scraper, Office 365 Activity, Office 365 Services, and Office 365 Message Trace. This data will not flow through Routes, but **will** be sent to the default Destination(s).

Workarounds: 1. In Collectors' config modals, open the **Result Routing** tab, Disable the **Send to Routes** default, and directly specify a **Pipeline** and **Destination**. (This option is not available in Prometheus Scraper or in the Office 365 Sources.) 2. Skip v.3.2.0.

Fix: In LogStream 3.2.1.

2021-11-17 – v.2.1.0 through Current – Re-enabling a Function group mistakenly re-enables all its Functions [CRIBL-7053]

Problem: When you change a Function group from disabled to enabled, all of its Functions are enabled, regardless of their individual enabled/disabled states when the group was disabled.

Workaround: Avoid disabling and re-enabling Functions as a group (e.g., for testing or stepwise debugging purposes).

Fix: Version TBD.

2021-11-17 – v.3.2.0 – TLS certs that use passphrases won't decrypt private keys [CRIBL-7049]

Problem: Sources whose TLS config uses a (known good) passphrase will fail to decrypt private keys. You will see a connect error, or an error of the form: `TLS validation error, is passphrase correct?`

Workarounds: 1. Edit the Leader's or single instance's `inputs.yml` file to insert a plaintext TLS passphrase. (See paths here.) Then commit and deploy the new config (distributed mode), or reload or restart the LogStream server (single instance). 2. Use a cert and key that do not require a passphrase. 3. Skip v.3.2.0.

Fix: In LogStream 3.2.1.

2021-11-17 – v.3.2.0 – Only one Chain Function works per Pipeline [CRIBL-7044]

Problem: If you add more than one Chain Function to a Pipeline, only the first will take effect. Chain Functions lower in the stack will simply pass the data down to the next Function.

Workaround: Design your data flow to require at most one Chain Function per Pipeline.

Fix: In 3.2.1.

2021-11-17 – v.3.2.0 – Exporting a Pack to another Group requires a Leader restart [CRIBL-7043]

Problem: Exporting a Pack to a different Worker Group via the UI succeeds, but opening the Pack on the target Group fails with a `Cannot read property...undefined` error.

Workaround: To resolve the error and make the target Pack accessible, restart the Leader. To prevent the error, export and import the Pack as a file.

Fix: In LogStream 3.2.1.

2021-11-16 – v.3.2.0 – QuickConnected Source goes to wrong Destination [CRIBL-7013, CRIBL-7047]

Problem: Some QuickConnect connections send data to an unintended Destination. We've observed this in single-instance deployments that include the Cribl-supplied `cribl_metrics_rollup` Pipeline, or include other Pipelines/Packs with stateful Functions like Rollup Metrics or Aggregations. Data will either flow through Routes instead of your specified QuickConnect Destination, or will continue flowing to the original QuickConnect Destination after you drag the connection to a different Destination.

Workaround: Use the **Data Routes** interface to manage the Pipeline and stateful Functions indicated above. If your QuickConnect data doesn't oblige a changed QuickConnect Destination, restart LogStream. This will stop data flow to the unintended Destination, and redirect it to the intended Destination.

Fix: In Cribl Edge 3.2.1.

2021-11-10 – v.3.2.0 – GitOps **Push** workflow displayed unsupported Revert button [CRIBL-6961]

Problem: When a GitOps `Push` workflow has placed the UI in read-only mode, the Commit UI displayed a **Revert** button, even though reverting changes was not supported. Pressing the button simply triggered an error message.

Fix: In Cribl Edge/LogStream 3.2.1.

2021-10-22 – v.3.2.x through Current – Kafka with Kerberos causes Workers' continuous restart [CRIBL-6674]

Problem: Configuring the Kafka Source or Destination with Kerberos authentication can send LogStream Workers into a continuous loop of restarts.

Workaround (multi-step):
1. In `krb5.conf`, set `dns_lookup_realm = false` and `dns_lookup_kdc = false`.
2. From `krb5.conf`'s `[realms]` section, extract the realm name (from the element name) and the FQDN (from the `kdc` key's value, stripping any port number).
3. Run `nslookup` on either or both of the realm name and the FQDN. E.g.:
`nslookup mysubdomain.confluent.io` and `nslookup kdc.kerberos-123.local`.
4. Add at least one of the returned IP addresses (which might be identical) to the `etc/hosts` file, as values in the format your platform expects.

Fix: Version TBD.

2021-10-11 – v.2.x-3.2.1 – Collectors do not filter correctly by date/time range [CRIBL-6440]

Problem: Collectors can retrieve data from undesired time ranges, instead of from the range specified. If paths are organized by date/time, this can also cause Collectors to retrieve data from the wrong paths. The root cause is that in the Collector's Run and Schedule configuration modals, time ranges are set based on the browser's time zone, whereas LogStream's backend assumes UTC.

Workaround: Offset the **Earliest** and **Latest** date/time values based on your browser's offset from UTC.

Fix: In Cribl Edge/LogStream 3.2.2 and later, the Run and Schedule modals provide a **Range Timezone** drop-down to prevent these errors.

2021-10-06 – v.3.1.2 – Monitoring page omits Collector Sources' data [CRIBL-6412]

Problem: With functioning Office 365 Activity and Office 365 Services Sources, the Job Inspector reports data being retrieved, and **Monitoring** > **Data** > **Destinations** reports data being sent out. However, **Monitoring** > **Data** > **Sources** falsely reports no data being received. The problem is isolated to this **Sources** page. It might also affect the Office 365 Message Trace and Prometheus Scraper Sources.

Workaround: Use the Source modal's **Live Data** tab, the **Monitoring** > **System** > **Job Inspector** page, and/or the **Monitoring** > **Data** > **Destinations** page to monitor throughput.

Fix: In LogStream 3.1.3.

2021-09-30 – v.3.1.2 – High CPU load with CriblMetrics Source [CRIBL-6319]

Problem: Enabling the CriblMetrics Source causes high event count and high CPU load.

Workaround: Adjust the `cribl_metrics_rollup` pre-processing Pipeline to roll up a wider **Time Window** of metrics.

Fix: Upgrade to LogStream 3.2.1, which provides an option to disable the CriblMetrics Source's **Full Fidelity** toggle.

2021-09-29 – v.3.0.2-3.1.2 – Memory leak with multiple Collectors [CRIBL-6310]

Problem: Configuring multiple Collectors can lead to gradual but cumulative memory leaks. Due to a caching error, the memory can be recovered only by restarting LogStream.

Workaround: Restart LogStream on the affected Worker Nodes.

Fix: In LogStream 3.1.3.

2021-09-21 – v.3.1.1 – Azure Blob Storage Source/Destination authentication error on upgrade from v.3.0.4 [CRIBL-6236]

Problem: Upon upgrading the Azure Blob Source and/or Destination from v.3.0.4 to v.3.1.1, you might receive an error of the form: `Unable to extract accountName with provided information.`

Workaround: Change the key, and then reset it back to your desired connection string.

Fix: Not observed as of Cribl LogStream 3.1.1.

2021-09-15 – v.3.0.0–3.1.3 – High CPU usage with Google Cloud Pub/Sub Source [CRIBL-6182]

Problem: The Google Cloud Pub/Sub Source substantially increases CPU usage, which stays high even after data stops flowing. This causes throughput degradation, and more-frequent `failed to acknowledge` errors.

Workaround: Configure the Google Cloud Pub/Sub Source's **Advanced Settings** > **Max backlog** to `1000`.

Fix: In 3.2.0, which defaults the **Max backlog** to `1000`, and also relaxes the retry interval from 10 seconds to 30 seconds.

2021-09-14 – v.3.1.1 – Modifying Collector > Preview > Capture settings can break Capture elsewhere [CRIBL-

6166]

Problem: Modifying the capture settings in a Collector's Run > Preview > Capture modal can improperly modify the Filter expression in Capture modals for other Collectors, Sources, and Routes/Pipelines.

Fix: In LogStream 3.1.2.

2021-09-10 – v.3.1.1 – Worker Group certificate name drop-down shows certificates from the Leader [CRIBL-6142]

Problem: When configuring certificates at **Groups** > `<group-name>` > **Settings** > **API Server Settings** > **TLS**, certificates configured on the Leader incorrectly appear on the **Certificate name** drop-down.

Fix: In LogStream 3.1.2.

2021-09-10 – v.3.1.1 – Git Collapsed Actions broken in 3.1.1 [CRIBL-6134]

Problem: With **Collapsed Actions** enabled, clicking the **Commit & Push** button has no effect. (The **Commit & Deploy** button works properly.)

Workaround: Disable **Collapsed Actions**, to restore separate **Commit** and **Git Push** buttons.

Fix: In LogStream 3.1.2.

2021-09-08 – All versions through v.3.1.1 – UDP support is currently IPv4-only [CRIBL-6106, CRIBL-6115]

Problem: Where Sources and Destinations connect over UDP, they currently support IPv4 only, not IPv6. This applies to Syslog, Metrics, and SNMP Trap Sources; and to Syslog, SNMP Trap, StatsD, StatsD Extended, and Graphite Destinations.

Workaround: Integrate via IPv4 if possible.

Fix: UDP Sources and Destinations gained IPv6 support in LogStream 3.1.2.

2021-09-02 – v.3.1.0-3.1.1 – Google Pub/Sub authentication via proxy environment variable fails [CRIBL-6086]

Problem: When LogStream's Google Cloud Pub/Sub Source and Destination attempt authentication through a proxy, using the `https_proxy` environment variable, they send an HTTP request to

http://www.googleapis.com:443/oauth2/v4/token. This request fails with 504/502 errors. The root cause is a mismatch in dependency libraries, whose correction has been identified, but requires broader testing.

Workaround: Configure the proxy in `transparent` mode, to avoid relying on environment variables.

Fix: In 3.1.2.

2021-08-24 – v.3.1.0 – High CPU load with LogStream version 3.1.0 [CRIBL-6039, CRIBL-6044]

Problem: LogStream 3.1.0 added code-execution safeguards that inadvertently increased CPU load, and decreased throughput, with several Functions and most expressions.

Workaround: Downgrade to version 3.0.4.

Fix: Upgrade to version 3.1.1 or later.

2021-08-18 – v.3.1.0 – Clone Collector option broken [CRIBL-5977]

Problem: Clicking a 3.1.0 Collector modal's **Clone Collector** button simply closes the modal. (If you have unsaved changes, you'll first be challenged to confirm closing the parent modal – but the expected cloned modal won't open.)

Workaround: Click **+ Add New** to re-create your original Collector's config from scratch, adding any desired modifications.

Fix: In LogStream 3.1.1.

2021-08-18 – All versions through 3.1.0 – Tabbed code blocks broken on in-app docs [CRIBL-5972]

Problem: When docs with tabbed code blocks are opened in the Help drawer, the default (leftmost) tab seizes focus. Other tabs will not display when clicked.

Workaround: Click the blue/linked page title atop the Help drawer to open the same page on docs.cribl.io, where all tabs can be selected.

Fix: In LogStream 3.1.1.

2021-08-14 – v.3.1.0 – Splunk Load Balanced Destination does not migrate auth type [CRIBL-5940]

Problem: In a Splunk Load Balanced Destination with **Indexer discovery** enabled and a corresponding **Auth token** defined, upgrading to LogStream 3.1.0 corrupts the **Auth token** field's value.

Workaround: Set the **Authentication method** to **Manual** and resave the token's value.

Fix: In 3.1.1.

2021-08-11 – v.3.1.0 – `C.Secret()` values are undefined in Collectors [CRIBL-5926]

Problem: Calling the C.Secret() internal method within a Collector field resolves incorrectly to an `undefined` substring. E.g., in URL fields, `C.Secret()` values will resolve to `/undefined/` path substrings.

Workarounds: 1. Use `C.vars` and a Global Variable, instead of using this method. 2. Root cause is that `C.Secret()` in Collectors and Pipeline Functions has access only to secrets that were created before the last restart. Therefore, restart Worker Processes to refresh the method's access.

Fix: In 3.1.1.

2021-08-10 – v.3.1.0 – Pre-processing Pipelines break Flows display [CRIBL-5909]

Problem: Attaching a pre-processing Pipeline to a Source breaks the **Monitoring** > **Flows (beta)** page's display. Attempting to remove Sources/Destinations from that page's selectors throws a cryptic `Sankey` error.

Workaround: Temporarily detach pre-processing Pipelines if you want to check Flows.

Fix: Upgrade to version 3.1.1 or later.

2021-07-29 – v.3.0.2-3.0.4 – Upgrades via UI require broader permissions [CRIBL-5774]

Problem: Upgrading from v.3.0.x via the UI requires the `cribl` user to be granted write permission on the parent directory above `$CRIBL_HOME`. The symptom is an error message of the form: `Upgrade failed: EACCES: permission denied, mkdir '/opt/unpack.xxxxxxx.tmp'`.

Workaround: Either adjust permissions, or upgrade via the filesystem. For complete instructions, see Upgrading.

Fix: Does not affect LogStream 3.1 or higher.

2021-07-26 – v.3.0.x–3.1.0 – Packs with orphaned lookups block access to Worker Groups [CRIBL-5738]

Problem: If a Pack references a lookup file that's missing from the Pack, pushing the Pack to a Worker Group will block access to the Group's UI. You will see an error message of the form: "The Config Helper service is not available because a configuration file doesn't exist... Please fix it and restart LogStream."

Workaround: On the Leader Node, review the config helper logs (`$CRIBL_HOME/log/groups/<group>/*.log`) to see which references are broken. (In a single-instance deployment, see `$CRIBL_HOME/log/*.log`.) Then manually resolve these references in the Pack's configuration.

Fix: Upgrade to version 3.1.1 or later.

2021-07-20 – v.3.0.3-3.4.0 – Can't add Functions to a Pipeline named `config` [CRIBL-5706]

Problem: You cannot add Functions to a Pipeline if the Pipeline is named `config`, because this name conflicts with the reserved route for the **Create Pipeline** dialog.

Workarounds: Don'tcha name your Pipelines `config`.

Fix: In Cribl Edge 3.4.1.

2021-07-06 – All versions through Current – Duplicate Workers/Worker GUIDs [CRIBL-5611]

Problem: Multiple Workers have identical GUIDs. This creates problems in Monitoring, upgrading and versioning, etc., because all Workers show up as one.

Cause: This is caused by configuring one Worker and then copying its `cribl/` directory to other Workers, to quickly bootstrap a deployment.

Workaround: Don't do this! Instead, use the Bootstrap Workers from Leader endpoint.

Fix: Version TBD.

2021-07-02 – v.3.0.2–3.0.3 – Sample file's last line not displayed upon upload [CRIBL-5595]

Problem: When uploading (attaching) a sample data file, the file's final line is not displayed in the **Add Sample Data** modal.

Workarounds: This is a UI bug only. LogStream correctly processes the complete sample data, which should show up when viewing the sample afterwards (e.g., within a Pipeline's preview pane).

Fix: In LogStream 3.1.0.

2021-07-02 – All versions through 3.0.x – Date fields misleadingly preview with string symbol [CRIBL-5594]

Problem: In Preview or Capture, incoming events like `_raw` will be displayed in the right pane with an `α` symbol that indicates string data. However, calling `new Date()` and then `C.Time.strptime()` methods in an Eval Function will return `null` on the OUT tab.

Cause: Due to the nature of JSON serialization, the incoming event's `Date` field is misleadingly subsumed under the event's `α` string symbol. It's actually a structured type, not a string…yet.

Workaround: If you see unexpected `null` results, stringify the datetime field as you extract it, e.g.: `new Date().toISOString()`. Feeding the resulting field to Time methods should return datetime strings as expected.

Fix: In LogStream 3.1.0.

2021-06-22 – v.3.0.0–3.0.3 – Internal `C.Text.relativeEntropy()` method – broken typeahead and preview [CRIBL-5534]

Problem: The `C.Text.relativeEntropy()` internal method is missing from JavaScript expressions' typeahead drop-downs. You can manually type or paste in the method, and save your Function and Pipeline, but LogStream's right Preview pane will (misleadingly) always show the method returning `0`.

Workarounds: Use other means (such as the **Live** button) to preview and verify that the method is (in fact) returning valid results.

Fix: In LogStream 3.1.0.

2021-05-20 – v.3.0.0 – Multiple Functions Break LogStream 3.0 Pipelines [CRIBL-5311, CRIBL-5766]

Problem: After upgrade to LogStream 3.0.0, including any of the following Functions in a Pipeline can break the Pipeline: GeoIP, Redis, DNS Lookup, Reverse DNS, Tee. Symptom is an error of the form: `Pipeline process timeout has occurred.` Less seriously, including these Functions in a Pipeline can suppress Preview's display of fields/values.

Workarounds: If you use these Functions in your Pipelines, stay with (or restore) a pre-3.0 version until LogStream 3.0.1 is available.

Fix: CRIBL-5311 fixed in LogStream 3.0.1. CRIBL-5766 fixed in LogStream 3.2.1.

2021-05-19 – v.3.0.0 – Leader's Changes fly-out stays open after Commit

Problem: In the Leader's left nav, the Changes fly-out remains stuck open after you commit pending changes.

Workarounds: Hover or click away. Then hover or click back to reopen the fly-out.

Fix: In LogStream 3.0.1.

2021-05-18 – v.3.0.0 – Packs > Export in "Merge" mode omits schemas and custom Functions

Problem: Exporting a Pack with the export mode set to `Merge` omits schemas and custom Functions configured within the Pack's **Knowledge** > **Schemas**.

Workarounds: 1. Change the export mode to `Merge safe`, and export again. 2. If that doesn't preserve the schema and Functions, revert to `Merge` export mode; install the resulting Pack onto its target(s); and then manually copy/paste the schema(s) and Functions from the source Pack's UI to the target Pack's UI.

Fix: In LogStream 3.0.1.

2021-05-17 – v.3.0.0 through Current – Can't Enable KMS on Worker Group after installing license

Problem: Enabling HashiCorp Vault or AWS KMS on a Worker Group, after installing a LogStream license on the same Group, fails with a spurious `External KMS is prohibited by the current license` error message.

Workaround: On the Leader, navigate to **Settings** > **Worker Processes**. Restart the affected Worker Group's `CONFIG_HELPER` process. Then return to that Worker Group's **Security** > **KMS** Settings, re-enter the same KMS configuration, and save.

Fix: Version TBD.

2021-05-10 – v.2.4.5-3.0.3 – Elasticsearch Destination, with Auto version discovery, doesn't send Authorization header

Problem: When the Elasticsearch Destination has Basic Authentication enabled, and its **Elastic version** field specifies `Auto` version discovery, LogStream fails to send the configured username and password credentials along with its API initial request. Elasticsearch responds with an HTTP 401 error.

Workaround: Explicitly set the **Elastic version** to either `7.x` or `6.x` (depending on your Elasticsearch cluster's version); then stop and restart LogStream to pick up this configuration change.

Fix: In LogStream 3.1.0.

2021-05-04 – v.2.4.5-3.0.1 – Office 365 Message Trace Source skips events

Problem: The Event Breaker Rule provided for the Office 365 Message Trace Source mistakenly presets the **Default timezone** to `ETC/GMT-0`. This setting causes LogStream to discover events but not collect them.

Workaround: Reset the Rule's **Default timezone** to `UTC`, then click **OK** and resave the Ruleset.

Fix: In 3.0.2.

2021-05-03 – v.2.4.4–3.0.1 – Rollup Function suppresses `sourcetype` metrics

Problem: `sourcetype` metrics can be suppressed when the Cribl Internal > CriblMetrics Source is enabled and the `cribl_metrics_rollup` pre-processing Pipeline is attached to a Source.

Workarounds: Disabling the pre-processing pipeline restores `sourcetype` and any other missing data. However, without the rollup, a much higher data volume will be sent to the indexing tier.

Fix: In LogStream 3.0.2.

2021-04-20 – v.2.4.3–2.4.5 – Orphaned S3 staging directories

Problem: Using the S3 Destination, defining a partitioning expression with high cardinality can proliferate a large number (up to millions) of empty directories. This is because LogStream cleans up staged files, but not staging directories.

Workaround: Programmatically or manually delete stale staging directories (e.g., those older than 30 days).

Fix: In LogStream 3.0.2.

2021-04-12 – v.2.4.4-3.0.0 – Splunk Sources do not support multiple-metric events

Problem: LogStream's Splunk Sources do not support multiple-measurement metric data points. (LogStream's Splunk Load Balanced Destination does.)

Fix: In LogStream 3.0.1.

2021-04-07 – v.2.4.2–2.4.5 – Google Cloud Storage Destination fails to upload files > 5 MB

Problem: The Google Cloud Storage Destination might fail to put objects into GCS buckets. This happens with files larger than 5 MB, and causes the Google Cloud API to report a vague `Invalid argument` error.

Workaround: Set the **Max file size (MB)** to 5 MB. Also, reduce the **Max file open time (sec)** limit from its default `300` (5 minutes) to a shorter interval, to prevent files from growing to the 5 MB threshold. (Tune this limit based on your observed rate of traffic flow through the Destination.)

Fix: In LogStream 3.0.0.

2021-03-31 – v.2.4.4-2.4.5 – Local login option visible even when disabled

Problem: The **Log in with local user** option is displayed to users even when you have disabled **Settings** > **Authentication** > **Allow local auth** for an OpenID Connect identity provider.

Workaround: Advise users to ignore this button. Although visible, it will not function.

Fix: In LogStream 3.0.0.

2021-03-31 – v.2.4.0–2.4.4 – Splunk TCP and LB Destinations' Workers trigger OOM errors and restart

Problem: With a Splunk TCP or Splunk Load Balanced Destination created after upgrading to LogStream 2.4.x, Workers' memory consumption may grow without bound, leading to out-of-memory errors. The API Process will restart the Workers, but there might be temporary outages.

Workaround: Toggle the Destination's **Advanced Settings** > **Minimize in-flight data loss** slider to `No`. This will preserve Processes killed by OOM conditions.

Fix: In LogStream 2.4.5.

2021-03-31 – v.2.4.4-2.4.5 – OpenID Connect authentication always shows local-auth fallback

Problem: Even if OpenID Connect external authentication is configured to disable **Allow local auth**, LogStream's login page displays a **Log in with local user** button.

Workaround: Do not click that button.

Fix: In LogStream 3.0.0.

2021-03-31 – v.2.4.4 – Authentication options mistakenly display Cribl Cloud

Problem: The **Settings** > **Authentication** > **Type** drop-down offers a `Cribl Cloud` option, which is not currently functional. Attempting to configure and save this option could lock the `admin` user out of LogStream.

Workaround: Do not select, configure, or save that option.

Fix: In LogStream 2.4.5.

2021-03-30 – v.2.4.4 – Can't disable some Sources from within their config modals

Problem: In configuration modals for the Azure Blob Storage and Office 365 Message Trace Sources, the **Enabled** slider cannot be toggled off, and its tooltip doesn't appear.

Workaround: Disable your configured Source (where required) from the **Manage Blob Storage Sources** or the **Manage Message Trace Sources** page.

Fix: In LogStream 2.4.5.

2021-03-29 – v.2.4.x – SpaceOut Destination is broken

Problem: Within the SpaceOut game, you cannot shoot, and your player is immortal.

Workaround: There are other video games. After we defeat COVID, you'll even be able to buy a PS5.

Fix: Restored in LogStream 2.4.5.

2021-03-24 – v.2.4.x – Cribl App for Splunk blocks admin password changes, configuration changes, and Splunk-based authentication

Problem: Attempting to change the admin password via the UI triggers a 403/Forbidden message. You can reset the password by editing `users.json`, but can't save configuration changes to Settings, Pipelines, etc., because RBAC Roles are not properly applied.

Workaround: Using a 2.3.x version of the App enables **local** authentication and enables changes to Cribl/LogStream passwords and configuration/settings.

Fix: In LogStream 2.4.4.

2021-03-22 – v.1.7.0-2.4.3 – Azure Event Hubs Destination: Compression must be manually disabled

Problem: LogStream's Azure Event Hubs Destination provides a **Compression** option that defaults to `Gzip`. However, compressed Kafka messages are not yet supported on Azure Event Hubs.

Workaround: Manually reset **Compression** to `None`, then resave Azure Event Hubs Destinations.

Fix: In LogStream 2.4.4.

2021-03-17 – v.2.4.2-2.4.3 – Parser Function > List of Fields copy/paste fails

Problem: When copying/pasting **List of Fields** contents between Parser Functions via the Copy button, the paste operation inserts unintended metadata instead of the original field references.

Workaround: Manually re-enter the second Parser Function's **List of Fields**.

Fix: In LogStream 2.4.4.

2021-03-13 – v.2.4.3 – UI can't find valid TLS .key files, blocking Master restarts and Worker reconfiguration

Problem: After upgrading to v.2.4.3, the UI fails to recognize valid TLS `.key` files, displaying spurious error messages of the form: "File does not exist: `$CRIBL_HOME/local/cribl/auth/certs/<keyname>key`." An affected Master will not restart. Affected Workers will restart, but will not apply changes made through the UI.

Workaround: Ideally, specify an absolute path to each key file, rather than relying on environment variables. If you're locked out of the UI, you'll need to manually edit the referenced paths within these configuration files in LogStream subdirectories: `local/cribl/cribl.yml` (General > API Server TLS settings) and/or `local/_system/instance.yml` (Distributed > TLS settings). Contact Cribl Support if you need assistance. A more drastic workaround is to disable TLS for the affected connections.

Fix: In LogStream 2.4.4.

2021-03-12 – v.2.4.2-2.4.5 – Redis Function with specific username can't authenticate against Redis 6.x ACLs

Problem: The Redis Function, when used with a specific username and Redis 6.x's Access Control List feature, fails due to authentication problems.

Workaround: In the Function's **Redis URL** field, point to the Redis `default` account, either with a password (e.g., `redis://default:Password1@192.168.1.20:6379`) or with no password (redis://192.168.1.20:6379). Do not specify a user other than `default`.

Fix: In LogStream 3.0.

2021-03-09 – v.2.4.3 – Splunk Destinations' in-app docs mismatch UI's current field order

Problem: For the Splunk Single Instance and Splunk Load Balanced Destinations, the in-app documentation omits the UI's **Advanced Settings** section. Some fields are documented out-of-sequence, or are omitted.

Workaround: Refer to the UI's tooltips, to the corrected Splunk Single Instance and Splunk Load Balanced online docs, and/or to the corrected PDF.

Fix: In LogStream 2.4.4.

2021-03-08 – v.2.4.3 – Enabling Git Collapse Actions breaks Commit & Deploy

Problem: After enabling **Settings** > **Distributed Settings** > **Git Settings** > **General** > **Collapse Actions**, selecting **Commit & Deploy** throws a 500 error.

Workaround: Disable the **Collapse Actions** setting, then commit and deploy separately.

Fix: In LogStream 2.4.4.

2021-03-08 – v.2.4.3 – S3 Collector lacks options to reuse HTTP connections and allow-self signed certs

Problem: As of v.2.4.3, LogStream's AWS-related Sources & Destinations provide options to reuse HTTP connections, and to establish TLS connections to servers with self-signed certificates. However, the S3 Collector does not yet provide these options.

Fix: In LogStream 2.4.4.

2021-03-04 – v.2.4.2 – Esc key closes both Event Breaker Ruleset modals

Problem: After adding a rule to a **Knowledge** > **Event Breaker Ruleset**, pressing `Esc` closes the parent Ruleset modal along with the child Rule modal.

Workaround: Close the Rule modal by clicking either its **Cancel** button or its close box.

Fix: In LogStream 2.4.3.

2021-03-04 – v.2.4.2 – Aggregations Function in post-processing Pipeline addresses wrong Destination

Problem: An Aggregations Function, when used in a post-processing Pipeline, sends data to LogStream's Default Destination rather than to the Pipeline's attached Destination.

Workaround: If applicable, use the Function in a processing or pre-processing Pipeline instead.

Fix: In LogStream 2.4.3.

2021-02-25 – v.2.4.2 – On Safari, Event Breaker shows no OUT events

Problem: When viewing an Event Breaker's results on Safari, no events are displayed on the Preview pane's **OUT** tab.

Workaround: Use another supported browser.

Fix: In LogStream 2.4.3.

2021-02-22 – v.2.4.3 – Collection jobs UI errors

Problem: Collection jobs are missing from the **Monitoring** > **Sources** page, even though they are returned by metric queries. Also, the **Job Inspector** > **Live** modal displays an empty, unintended **Configure** tab.

Workaround: Use the Job Inspector to access collection results. Ignore the **Configure** tab.

Fix: In LogStream 2.4.4.

2021-02-19 – v.2.4.2 – Upon upgrade, Git remote repo setting breaks, blocking Worker Groups

Problem: If a Git remote repo was previously configured, upgrading to LogStream v.2.4.2 throws errors of this form upon startup: `Failed to initialize git repository. Config versioning will not be available...Invalid URL...`. The Master cannot commit or deploy to any Worker Group.

Workarounds: 1. Downgrade back to v.2.4.1 (or your previous working version). 2. Switch from Basic authentication to SSH authentication against the repo, to remove the username from requests.
(The apparent root cause is Basic/http auth using a valid URL and username, but missing a password.)

Fix: In LogStream 2.4.3.

2021-02-19 – v.2.4.0-2.4.2 – Splunk (S2S) Forwarder access control blocks upon upgrade to LogStream 2.4.x

Problem: If Splunk indexers have forwarder tokens enabled, and worked with LogStream 2.3.x before, upgrading to LogStream 2.4.x causes data to stop flowing.

Workaround: If you encounter this problem, rolling back to your previously installed LogStream version (such as v.2.3.4) resolves it.

Fix: In LogStream 2.4.3.

2021-02-10 – v.2.4.0-2.4.1 – With Splunk HEC Source, JSON payloads containing embedded objects trigger high CPU usage

Problem: Splunk HEC JSON payloads containing nested objects trigger high CPU usage, due to a flaw in JSON parsing.

Workaround: If you encounter this problem, rolling back to your previously installed LogStream version (such as v.2.3.4) resolves it.

Fix: In LogStream 2.4.2.

2021-01-30 – v.2.4.0 – Worker Nodes cannot connect to Master

Problem: Worker Nodes cannot connect to the Master after the Master is upgraded to v.2.4.0.

Workaround: Disable compression on the Workers. You can do so through the Workers' UI at **System Settings** > **Distributed Settings** > **Master Settings** > **Compression**, or by commenting out this line in each Worker's `cribl.yml` config file:

```
compression: gzip
```

Fix: In LogStream 2.4.1.

2021-01-25 – v.2.4.0 – S3 collection stops working due to auth secret key issues.

Problem: S3 collection stops after upgrade to 2.4.0 due to secret key re-encryption.

Workaround: Re-configure S3, save and re-deploy.

Fix: In LogStream 2.4.1.

2021-01-14 – v.2.4.0 – Google Cloud Storage Destination Needs Extra Endpoint to Initialize

Problem: The Google Cloud Storage Destination fails to initialize, displaying an error of the form: `Bucket does not exist!`

Workaround: In the `outputs.yml` file, under your `cribl-gcp-bucket` key endpoint, add: `https://storage.googleapis.com`. (in a single-instance deployment, locate this file at `$CRIBL_HOME/local/cribl/outputs.yml`. In a distributed deployment, locate it at `$CRIBL_HOME/groups/<group name>/local/cribl/outputs.yml`.)

Fix: In LogStream 2.4.1.

2021-01-14 – v.2.4.0 – Worker Groups' Settings > Access Management Is Absent from UI

Problem: In this release, the **Worker Groups** > **<group-name>** > **System Settings** UI did not display the expected **Access Management**, **Authentication**, and **Local Users** sections.

Workaround: Manually edit the `users.json` file.

Fix: In LogStream 2.4.1.

2021-01-13 – v.2.4.0 – Route Filters Aren't Copied to Capture Modal

Problem: On the **Routes** page, selecting **Capture New** in the right pane does not copy custom **Filter** expressions to the resulting **Capture Sample Data** modal. That modal's **Filter Expression** field always defaults to `true`.

Workarounds: 1. Bypass the **Capture New** button. Instead, from the Route's own ••• (Options) menu, select **Capture**. This initiates a capture with the **Filter Expression** correctly populated. 2. Copy/paste the expression into the **Capture Sample Data** modal's **Filter Expression** field. Or, if the expression is displayed in that field's history drop-down, retrieve it.

Fix: In LogStream 2.4.1.

2021-01-13 – v.2.4.0 – Destinations' Documentation Doesn't Render from UI

Problem: Clicking the **Help**<img class="inline" src="/images/page/help-button.png" alt="Help link" width="
width="10%"" height="10%">link in a Destination's configuration modal displays the error message: "Unable to load docs. Please check LogStream's online documentation instead."

Workarounds: 1. Go directly to the online Destinations docs, starting here. 2. Follow the UI link to the docs landing page, click through to open or download the current PDF, and scroll to its Destinations section.

Fix: In LogStream 2.4.1.

2021-01-13 – v.2.4.0 – `Esc` Key Doesn't Consistently Close Modals

Problem: Pressing `Esc` with focus on a modal's drop-down or slider doesn't close the modal as expected. (Pressing `Esc` with focus on a free-text field, combo box, or nothing does close the modal – displaying a confirmation dialog first, if you have unsaved changes.)

Workarounds: Click the **X** close box at upper right, or click **Cancel** at lower right.

Fix: In LogStream 2.4.1.

2020-12-17 – v.2.3.0-2.4.0 – Free-License Expiration Notice, Blocked Inputs

Problem: LogStream reports an expired Free license, and blocks inputs, even though Free licenses in v.2.3.0 do not expire.

Workaround: This is caused by time-limited Free license key originally entered in a LogStream version prior to 2.3.0. Go to **Settings > Licensing**, click to select and expand your expired Free license, and click **Delete license**. LogStream will recognize the new, permanent Free license, and will restore throughput.

Fix: In LogStream 2.4.1.

2020-11-14 – v.2.3.3 – Null Fields Redacted in Preview, but Still Forwarded

Problem: Where event fields have null values, LogStream (by default) displays them as struck-out in the right Preview pane. The preview is misleading, because the events are still sent to the output.

Workaround: If you do want to prevent fields with null values from reaching the output, use an Eval Function, with an appropriate Filter expression, to remove them.

Fix: Preview corrected in LogStream 2.3.4.

2020-10-27 – v.2.3.2 – Cannot Name or Save New Event Breaker Rule

Problem: After clicking **Add Rule** in a new or existing Event Breaker Ruleset, the **Event Breaker Rule** modal's **Rule Name** field is disabled. Because **Rule Name** is mandatory field, this also disables saving the Rule via the **OK** button.

Fix: In LogStream 2.3.3.

2020-10-12 – v.2.3.1 – Deleting One Function Deletes Others in Same Group

Problem: After inserting a new Function into a group and saving the Pipeline, deleting the Function also deletes other Functions lower down in the same group.

Fix: In LogStream 2.3.2.

Workaround: Move the target Function out of the group, resave the Pipeline, and only then delete the Function.

2020-09-27 – v.2.3.1 – Enabling Boot Start as Different User Fails

Problem: When a root user tries to enable boot-start as a different user (e.g., using `/opt/cribl/bin/cribl boot-start enable -u <some-username>`), they receive an error of this form:

```
error: found user=0 as owner for path=/opt/cribl, expected uid=NaN.
Please make sure CRIBL_HOME and its contents are owned by the uid=NaN by running:
[sudo] chown -R NaN:[$group] /opt/cribl
```

Fix: In LogStream 2.3.2.

Workaround: Install LogStream 2.2.3 (which you can download here), then upgrade to 2.3.1.

2020-09-17 – v.2.3.0 – Worker Groups menu tab hidden after upgrade to LogStream 2.3.0

Problem: Upon upgrading an earlier, licensed LogStream installation to v.2.3.0, the **Worker Groups** tab might be absent from the Master Node's top menu.

Fix: In LogStream 2.3.1.

Workaround: Click the **Home** > **Worker Groups** tile to access Worker Groups.

2020-09-17 – v.2.3.0 – Cannot Start LogStream 2.3.0 on RHEL 6, RHEL 7

Problem: Upon upgrading to v.2.3.0, LogStream might fail to start on RHEL 6 or 7, with an error message of the following form. This occurs when the user running LogStream doesn't match the LogStream binary's owner. LogStream 2.3.0 applies a restrictive permissions check using `id -un <uid>`, which does not work with the version of `id` that ships with these RHEL releases.

```
id: 0: No such user
ERROR: Cannot run command because user=root with uid=0 does not own executable
```

Fix: In LogStream 2.3.1.

Workaround: Update your RHEL environment's `id` version, if possible.

2020-09-17 – v.2.3.0 – Cannot Start LogStream 2.3.0 with OpenId Connect

Problem: Upon upgrading an earlier LogStream installation to v.2.3.0, OIDC users might be unable to restart the LogStream server.

Fix: In LogStream 2.3.1.

Workaround: Edit `$CRIBL_HOME/default/cribl/cribl.yml` to add the following lines to its the `auth` section:

```
filter_type: email_whitelist
scope: openid profile email
```

2020-06-11 – v.2.1.x – Can't switch from Worker to Master Mode

Problem: In a Distributed deployment, attempting to switch Distributed Settings from Worker to Master Mode blocks with a spurious "Git not available…Please install and try again" error message.

Fix: In LogStream 2.3.0.

Workaround: To initialize `git`, switch first from Worker to Single mode, and then from Single to Master mode.

2020-05-19 – v.2.1.x – Login page blocks

Problem: Entering valid credentials on the login page (e.g., `http://localhost:9000/login`) yields only a spinner.

Fix: In LogStream 2.3.0.

Workaround: Trim `/login` from the URL.

2020-02-22 – v.2.1.x – Deleting resources in `default/`

Problem: In a Distributed deployment, deleting resources in `default/` causes them to reappear on restart.

Workaround/Fix: In progress.

2019-10-22 – v.2.0.0 – In-product upgrade issue on v2.0

Problem: Using in-product upgrade feature in v.1.7 (or earlier) fails to upgrade to v.2.0, due to package-name convention change.

Workaround/Fix: Download the new version and upgrade per steps laid out here.

2019-08-27 – v.1.7 – In-product upgrade issue on v.1.7

Problem: Using in-product upgrade feature in v.1.6 (or earlier) fails to upgrade to v.1.7 due to package name convention change.

Workaround/Fix: Download the new package and upgrade per steps laid out here.

2019-03-21 – v.1.4 – S3 stagePath issue on upgrade to v.1.4+

Problem: When upgrading from v.1.2 with a S3 output configured, `stagePath` was allowed to be undefined. In v.1.4+, `stagePath` is a required field. This might causing schema violations when upgrading older configs.

Workaround/Fix: Reconfigure the output with a valid `stagePath` filesystem path.

18.2. Common Challenges On The Edge

Here are some common issues that users have encountered when using Cribl Edge, along with guidance to resolve them.

Licensing and Quotas Confusion

Managed Edge Nodes use their Leader's licensed ingestion quota, but only for Sources that are enabled and configured. Data from some Sources like system state that collect data locally are not counted against the license. Data sent to Cribl Stream via TCP/HTTP is not counted against the license. For details see, Licensing.

4.3.0 Upgrading Errors

If a Leader is on version 4.3.0, you should upgrade all Edge Nodes to 4.3.0 before making any config changes or committing and deploying to the Nodes.

If you have already committed and deployed to Nodes that are on an earlier version (and the Leader is on 4.3.0), you won't be able to upgrade the Edge Nodes.

In such a scenario, you might see errors like these:


```
{
"time": "2023-09-14T00:38:34.919Z",
"cid": "api",
"channel": "rest:master-workers",
"level": "error",
"message": "API Error",
"error": {
"message": "Item not found",
"stack": "RESTError: Item not found\n at b.throwError (/opt/cribl/bin/cribl.js:14:1
},
"

},
"url": "/master/workers/6f822de1-d739-45ae-b29d-097d6e1cb647"
}
```


For workarounds, see Upgrading to 4.3.0 using Leader-Managed Upgrades for Edge Nodes.

Web UI Errors

While we hope you have an error-free experience, we want to prepare you in case you run into issues. Here are some errors you might encounter in Edge, and how to resolve them:

Error: “No workers registered”

Cause: Occurs during a live data capture, when there are no Edge Nodes available to fulfill the capture request. This error can be triggered if you attempt to capture events too quickly following a commit & deploy.

Recommendation: There are two things you can try:

- If you recently committed and deployed changes, wait 30 seconds and retry the live capture.
- Navigate to **Manage** > **Edge Nodes** and confirm there are Nodes currently registered with the Leader. Note that the error appears on a per-Fleet basis, which means that even if you have registered Nodes, they may not be in the Fleet you’re capturing live data from.

Leader and Edge Node (In)Compatibility

Leaders on v.4.2.x are compatible with Edge Nodes on v.4.1.2 and later. Due to a security update, Edge Nodes running on v.4.0.4 cannot receive configurations from v.4.2.x Leaders. For details, see Leader and Edge Nodes Compatibility.

Duplicate Edge Node GUID

When you install and first run the software, a GUID is generated and stored in a `.dat` file located in `CRIBL_HOME/bin/`. For example:

```
# cat CRIBL_HOME/bin/676f6174733432.dat

{"it":1570724418,"phf":0,"guid":"48f7b21a-0c03-45e0-a699-01e0b7a1e061"}
```

When deploying Cribl Edge as part of a host image or VM, be sure to remove this file so that you don’t end up with duplicate GUIDs. The file will regenerate on the next run.

Running Edge (Linux) as Non-root User

Privileged access might be necessary if Cribl Edge needs to read certain resources (e.g., `/var/log/*`), or to listen on low ports `1–1024`. Features like auto-discovery of logs and information in the Processes UI also require permissions to access `/proc`. The regular non-root permissions are not sufficient in these cases. For guidelines, see Running Edge as an Unprivileged User.

Debugging Cribl Edge on Windows

Common issues users encounter when deploying Cribl Edge on Windows include:

Using a Pre-4.3.0 MSI Installer

MSIs earlier than v.4.3.0 install the app so it belongs to the user running the installer. If you later try to upgrade it as another user, things go awry — resulting in a rather confusing mess with a seemingly random mix of the old and new package contents on disk.

Moving forward, we recommend adding `ALLUSERS=1` to the `msiexec` command when installing or upgrading pre-4.3.0 versions for testing. Example command:

```
msiexec /i cribl-<version>-<build>.msi /qn MODE=mode-managed-edge HOSTNAME=
<yourhostname> PORT=4200 AUTH=myAuthToken FLEET=myfleet USERNAME=LocalSystem
ALLUSERS=1
```

Debugging MSI Installations

To debug MSI installation or upgrade issues, you can run `msiexec.exe` to set logging options. See Microsoft's Logging Options topic.

For troubleshooting an MSI installation that does not pull configuration settings from the Leader node, you need to know the log path on Windows: The full path depends on where you chose to install.

`$INSTALL_PATH\cribl\log` and `local\edge\` are the log location and conf location, respectively.

CPU/Memory Issues on Windows

If you notice high CPU/Memory usage on your Cribl Edge Windows:

- The Users and Groups collector in the System State source can trigger high CPU utilization in the Windows lsass.exe process where there are many user accounts on the host; this is not uncommon on Domain Controllers. Consider disabling that collector if you see that process consuming excessive CPU.
- If you are interested in collecting logs only, disable the Windows Metrics source.

- During the Windows Event Log capture, run the collection for 10 seconds and tweak the settings if necessary so that collection doesn't always have to catch up. Consider changing the Windows Events Log Batch Size and Frequency.
  Example command to capture a certain number of events:
  ```
  measure-command {Get-WinEvent -Oldest -MaxEvents 500 @{LogName='Security'} | ForEach-Object -Process {ConvertTo-Json -Compress $_}}
  ```
- For Domain Controllers with heavy log volumes, consider using Windows Event Forwarder and send them to Stream instead of using Windows Event Log Source in Cribl Edge.

Local Files When Uninstalling Windows

This uninstall process does not automatically delete the local, log, pid, and state directories, because those folders include contents generated after the original installation completed. To completely remove the application, you must explicitly delete these folders from the filesystem.

Edge via Kubernetes

Common mistakes users make when deploying via Kubernetes include:

Failing Health Checks

By default, the Cribl Edge API listens on port 9420, instead of on Cribl Stream's default 9000 port. This requires special care when you configure your Edge Fleet in the Cribl UI.

The API server is configured to listen on the loopback interface (127.0.0.1). Change the listening interface address to 0.0.0.0 to bind to all IPs in the container. If you do not make this change, your health checks will start failing after the Edge agent downloads its initial config bundle.

Another side effect of not changing the Listening Interface address, is the Edge Node will not appear on the Leader UI.

Not Setting the Environment Variables for Invalid Certifications

If your Kubernetes deployment doesn't use valid certificates, set the `CRIBL_K8S_TLS_REJECT_UNAUTHORIZED` environment variable to 0 to disable that expectation. When you disable this environment variable, all Kubernetes features (including Metadata, Metrics, Logs, and AppScope metadata) will tolerate invalid (expired, self-signed, etc.) TLS certificates when connecting to the Kubernetes APIs. If this environment variable is not defined or set to 1 and the certificate validation fails, all connection attempts to the Kubernetes API will be rejected.

Not Authorizing the Cribl Edge Pod Service Account

Cribl Edge's Kubernetes Logs and Kubernetes Metrics Sources use the Kubernetes API to read logs and metrics from the cluster. If you have role-based access control (RBAC) implemented, you will need to authorize the Cribl Edge Pod Service Account to read the relevant data. For details, see RBAC.

Not Deploying Cribl Edge as a Daemonset

The Kubernetes Sources (Events, Logs, Metrics) can generate redundant events when it is running inside the cluster and can't determine which DaemonSet it's in or which Node it's on. This happens when you run Cribl Edge inside the cluster in a way that it can't detect the local Pod/Node. In these cases, Edge will emit an error when it initializes this Source.

You can bypass the error state by setting the CRIBL_K8S_FOOTGUN environment variable to true. However, you will assume the risk that the Kubernetes Events Source might make excessive calls to the cluster API.

For the Kubernetes Events Source to work as designed, we recommend deploying Cribl Edge as a Daemonset. For details, see Deploying via Kubernetes.

Which Destination to Use for Cribl Stream

Always use Cribl TCP or Cribl HTTP as the Destinations to send data to Cribl Stream. Otherwise, you might see duplication of certain data (such as host). Also, data sent to Cribl Stream via TCP/HTTP is not counted against the license.

Unsupported Files on the Journal Files Source

This Source does not support the new `COMPACT` binary file format. The logs will contain an error message when the parser encounters this type of file. (A fix is planned for v.4.3.1.). Sample error message when enabling the Journal Files Source:

```
{ [- ]
channel: input: in_journal_local cid: WO err: { [-]
message: Cannot read properties of undefined (reading 'data')
stack: TypeError: Cannot read properties of undefined (reading 'data')
at w. readEntry (/opt/cribl-edge/bin/crib1.js:14:18884713)
at w. readEvents (/opt/cribl-edge/bin/crib1.js:14:18885930)
at w._transform (/opt/crib1-edge/bin/crib1.js:14:18886631)
at w. Transform. _write (node: internal/streams/transform:205:23) at writeOrBuffer
}
level: warn
message: Error while stopping journal collector time: 2023-08-15T21:04:51.929Z
Show as raw text
host = ip-10-99-99-38.us-west-2.compute. internal
source = cribl: sourcetype = cribledge
```

Also, this Source does not currently support all of the compression options that the journald service supports. The parser this Source uses to read the files can handle `LZ4` and `ZSTD` compression. When the parser encounters compressed fields it can't unpack, an error message emits in the logs. Here is an example of the error message:

```
Error reading journald event, dropping offset=31678152 evt="Mar 02 21:29:54 goats-xps15 multipassd: undefined"
```

File Monitor Source

Common mistakes users make when exploring the File Monitor Source include the following:

Using an Editor

Do not test the Source by editing and saving files in a directory using a text editor (vi or Notepad). Under the hood, these editors operate on a temporary file that they move into place when saved. As a result, the File Monitor Source sees files get replaced and moved rather than appended to. This confuses the Source's logic that assumes logs are appended to the existing file.

Run the command below to append logs to files: `echo "{ \"time\":\"$(date -Is -u)\", \"message\":\"...\" }" >> eg.log`

You end up with lines like this in the logs: `{ "time":"2023-09-19T20:08:42+00:00", "message":"..." }`

Misunderstanding the Filename Allowlist

The Filename allowlist in the Source applies the wildcard patterns to the entire path of the discovered file, not just the file name.

The search path itself is not excluded from the match. For example, if you want to collect the file /var/log/messages when you've set your search path to `/var/log`, your filename allowlist needs to be either `/var/log/messages`, `/var/*/messages`, or `*/log/messages`.

For details, see Using Allowlists Effectively.

Not configuring the Hash Length to Avoid Common Headers

It's common for a batch of CSV files or IIS logs to have identical first lines listing columns that appear in subsequent lines. For files with identical first lines, configure the Hash length to be longer than the first line to ensure that the header includes something unique. For details, see Configuring Hash Lengths.

Allowing Race Conditions

The hashing logic isn't instantaneous, so it's possible for the common header problem described above to be confusing. Quickly copying the same example file to separate eg1.log and eg2.log files in the Source's path will usually result in both files being collected even though they share the same hash. However, copying it to eg3.log shortly (not immediately) after will result in that third file being ignored because its head-hash matches an existing entry in the state-store.

Testing Multiple Copies of the Same File

The head-hashing logic used by the Source to determine what it's seen before and how far it's collected already prevents you from successfully testing multiple copies of the same. You can't copy the same file into the Source's path over and over and expect them to be collected repeatedly.

Pointing Multiple File Monitor Sources to the Same File

Multiple File Monitor Sources set to collect the same file will result in duplicate events. Sources run independently and don't coordinate with each other.

Testing with Small Files

The Source doesn't track state using filenames. Instead, it uses hashes of the head and tail of the file to identify it later so it can resume where it left off. There is some exception logic for dealing with files smaller

than the hash sizes that cannot be made perfect. If you try to test with small files (smaller than 256 bytes), you will be working with the fallback logic and not the logic used for normal logs.

Panicking over Variable File Sizes

The File Monitor Source can handle a variety of file sizes without “starving” the smaller ones. In Cribl Edge, multiple workflows can be multiplexed using the event loop, even though it runs on a single core.

Not Testing with Event Breakers

Test sample file content with Event Breakers to make sure the incoming files are splitting into individual events as expected. When you configure the Event Breaker ruleset, configure the timestamp correctly and verify that it works. The Event Breaker will default to “Now” if it can’t find a timestamp in the raw log, and the cursor state is saved at that point.

If it encounters another log later in the file with a timestamp before “Now”, it will ignore it. For details, see Timestamp Settings.

18.3. Working With Cribl Support

If you run into issues with Cribl Edge, please first check our Known Issues page for recommended resolutions or workarounds. For questions not addressed there, this page outlines how to engage with the Cribl Support staff to resolve problems as quickly as possible.

Contacting Cribl Support

You can contact Cribl Support in multiple ways, outlined below:

- Email
- Community Slack
- Cribl Curious
- Support Portal (login required)

The best method depends on your reason for contacting Support. For quick, question-and-answer discussions, Community Slack or Cribl Curious might be sufficient. For in-depth discussions, or to troubleshoot technical issues, create a support case for better tracking and exchange of information.

Creating a Support Case

You have two options for opening a support case: via email or via the Support Portal (access here, details below).

If you exchange diags or other files with us, note that all file-transfer methods **except** for email are encrypted.

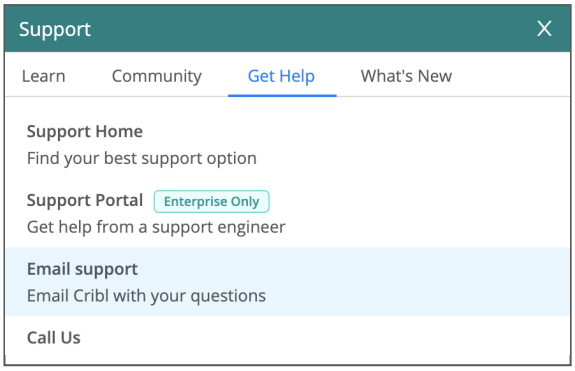
Email communication imposes a total message-size limit of 25 MB (after MIME-encoding attachments). The Support Portal allows file attachments of up to 100 MB each.

Contact via Email

The simplest way to engage with Support is to email support@cribl.io, with the information outlined below. This will automatically open a case for us to track, and will send you an auto-confirmation email that includes your case number. A Support engineer will then contact you to begin troubleshooting.

Email via Support Drawer

Within Cribl Edge's UI, you can click the top-right **Support** button and, in the resulting **Support** drawer, select the **Get Help** upper tab. Then click **Email support**. This will create a new email to us in your email client.



Email from the UI

Contact via Community Slack

Our Support staff monitors Cribl's Community Slack for any issues customers are experiencing: https://cribl-community.slack.com/. If you are not already registered for our Community Slack, please register at https://cribl.io/community/ to get started.

Slack might not get you the same timely response as an email, but it's a great way to get questions about Cribl Edge answered by a wide range of Cribl insiders, and expert peer users, who enjoy sharing their knowledge of the product. Check out individual channels dedicated to feature requests, docs, and other concerns.

Private Slack Channels

Our Enterprise customers have their own private channels where they can communicate directly with their Cribl account team.

Contact via Cribl Curious

Cribl Curious is a threaded community Q&A site. Our support staff monitors it, as do other Cribl teams. As with Community Slack, it's best for issues that aren't critical or urgent.

Contact via Support Portal

The Cribl Support Portal is available to our licensed customers. It facilitates encrypted transfer of large files, and maintains support cases' history so that you can easily review them.

The Support Portal uses single sign-on (SSO) through an integration with Cribl.Cloud accounts. So as part of Support Portal signup, you'll receive an invitation to create a Cribl.Cloud account, if you don't already have

one. You have no obligation to use your Cribl.Cloud Organization for purposes other than SSO, if it doesn't meet your needs.

For details about navigating Cribl.Cloud, see the Cribl.Cloud.

The Support Portal enables two types of users: Standard and Admin. You can have up to four user logins per customer account, one of which can be (but is not required to be) an Admin user.

Standard Versus Admin Users

Standard users can:

- Create cases.
- Manage cases.
- Search cases.
- Upload files to cases.
- Access other Cribl resources via their Cribl.Cloud portal.

Admin users can do all of the above, plus:

- View all cases on the customer's account.
- Edit case information, post-creation.
- Invite Standard users to the portal.

Signing Up

You'll need to receive a Cribl.Cloud invitation from Cribl or your Support Portal Admin, and follow the directions in the email to sign up. Both scenarios are summarized below. But first, a word from our sponsor, us:

Video Tutorial: Signing Up, Submitting Cases

This video walks you through signing up for the Support Portal, and then submitting cases with enough detail for Cribl Support to rapidly help you.

How to Access the Cribl Support Portal

06:01

Accessing the Cribl Support Portal

Standard User Signup

To access the Cribl Support Portal if you already have a Cribl.Cloud account:

1. Contact Cribl Support to request a user login to the Support Portal.
2. Cribl Support will send an invitation to you via email.
3. Follow the link in the email to sign up, using the same email address at which you received the invitation.
4. Log in with your existing Cribl.Cloud account.
5. Once logged in, you can create support cases, view any of your open or closed cases, etc.

To access the Cribl Support Portal without an existing Cribl.Cloud account:

1. Contact Cribl Support at support@cribl.io to request to be added. Cribl Support will email you an invitation. (Your account's Support Portal Admin has the capability to invite you too.)
2. Follow the link in the email to sign up, using the same email address at which you received the invitation.
3. Register your new Cribl.Cloud account.
4. Then log into https://portal.support.cribl.io with your new Cribl.Cloud account.
5. Once logged in, you can create support cases, view any of your open or closed cases, etc.

Admin User Signup

To access the Cribl Support Portal as an Admin user:

1. Contact Cribl Support to request Admin access.
2. If your customer account already has an Admin user, the current Admin must also contact Cribl, requesting to transfer the account's sole Admin user role to you.
   The request from the current Portal Admin must originate from the email address we have on file for that Admin user. We can tell you who the current Admin is, if you do not know.
3. If you have an existing Cribl.Cloud account, then Cribl Support will promote your account to Admin. If you do not yet have a Cribl.Cloud account, Cribl Support will email you a Support Portal signup invitation. Follow the link in the email to accept the invitation, using the same email address at which you received the invitation.

Submitting Cases via the Portal

All support cases must be submitted by an individual account (not a shared or group account). However, there are broader notification options.

When you create a new case, note the two fields for **Related Teammates**. Each name entered here must be an existing Contact on your customer account. If you want to enter one of your own organization's **group** email addresses, contact Cribl Support to create a contact entry for that address.

Relevant Information We Need

When contacting Cribl Support via any means, please provide as many of the following details as you can – the more, the better:

- Your name.
- Preferred contact method (phone or email).
- Cribl Edge version number affected.
- Description of the issue you're having.
- What's the issue's scope? (Leader Node, specific Edge Nodes or Fleets, or entire deployment; number of nodes impacted).
- When did the issue begin, or when was it first noticed?
- Did you make any known changes around that time? (Upgrade, config change, network change, etc.).
- Diags for one or more affected systems (the Leader Node does not process data, so typically, only diags from Workers are necessary).
- Sample event data for testing Pipeline issues (provide a text file, rather than screenshots).

- Any troubleshooting steps that you've already taken.

Pulling Diags

Providing us diags from your environment will speed up the time to resolution. For instructions on how to pull a diag file, see Diagnosing Issues.

With Cribl Stream (LogStream) 2.4.1 or later, and Cribl Edge, the diag is uploaded from the browser, rather than from the Node. This means that your Edge Nodes do not need internet access.

With LogStream 2.4.0 or earlier, you'll need to transfer the diags from your Worker Nodes.

> Diag bundles for Leader Nodes do not include diag bundles for any Edge Nodes.
>
> If you would like to upload your diag file via the GUI or CLI, you'll need outbound internet access to the upload URL (https://diag-upload.cribl.io), plus a valid support case number (provided in your case confirmation email).

Diag Workarounds

If your organization does not permit outbound access to https://diag-upload.cribl.io to upload from within Cribl Edge, you can also submit diags directly via the Support Portal (login required).

If none of these options work with your organization's policies, please work directly with your Support engineer to find a solution.