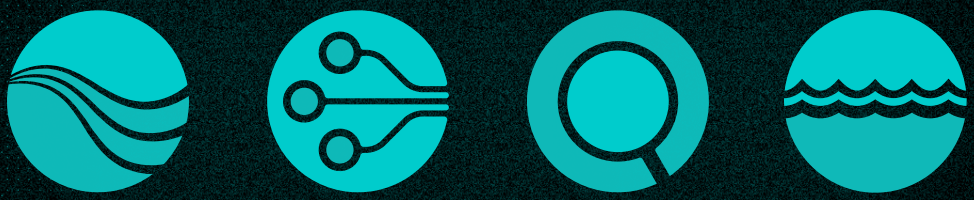
Cribl

Cribl Solution Guide



Cribl Solution Guide

Executive summary.

In today's digital era, businesses struggle to make the most of their data. They face numerous hurdles: scaling to keep pace with data growth, increasing cyberthreats, tight budgets, and a lack of talent. Because of these challenges, IT and Security teams often have to compromise on what data they deliver to which analysis tools.

But what if you could make choices without compromising? What if you could unlock the value of all of your data no matter how fast it grew and deliver it in any format to any tool that needs it?

Cribl's suite of products is designed to address the complex challenges of today's digital landscape, offering a solution that lets you use all your data without compromise. Cribl's innovative tools transform data management, so teams finally have the control, flexibility, and efficiency they need to make data work for them and their customers.

Document purpose:

This guide provides a holistic overview of how Cribl solutions can help your organization unlock the value of data, ensuring it's not just abundant but actionable and impactful. Read on for an overview of the solution's features, key capabilities, benefits, and deployment options.

Solution overview.

Introduction.

The Cribl suite of products is designed with a singular vision: to empower organizations to regain control of their data landscape. Our solutions let you choose what is best for your organization, the control to get the data where you want, and the flexibility to put it in any format you need on the fly.

Cribl Stream optimizes data in motion, enabling you to route, enrich, and transform data efficiently. Cribl Edge extends these capabilities to the edge, bringing processing closer to the source and reducing latency. Cribl Search revolutionizes data exploration, giving you real-time insights from stored and live data. Cribl Lake makes it easy for organizations to easily store, manage, and access data, so data can be usable and valuable to the teams and tools that need it.

Each product shows Cribl's commitment to giving you flexibility in data management, ensuring you have the right tools to make informed decisions and drive your business forward.

Solution architecture.

The figure below shows an overview of Cribl's solution architecture. Cribl Stream and Cribl Edge make it a breeze to collect, process, and deliver observability, security, or telemetry data in real time to wherever you need it. Use out-of-the-box, ad hoc data collection capabilities to get data from your lakes to your analytics tools, and perform search-in-place queries with Cribl Search to increase the scope of analysis – enabling compliance and insights. With Cribl Lake, you can easily share and route data downstream through Stream and Edge in any format, to any tool, at any time. Search unifies the query experience no matter where data is stored, so there's no need to move or rehydrate your data to analyze it.

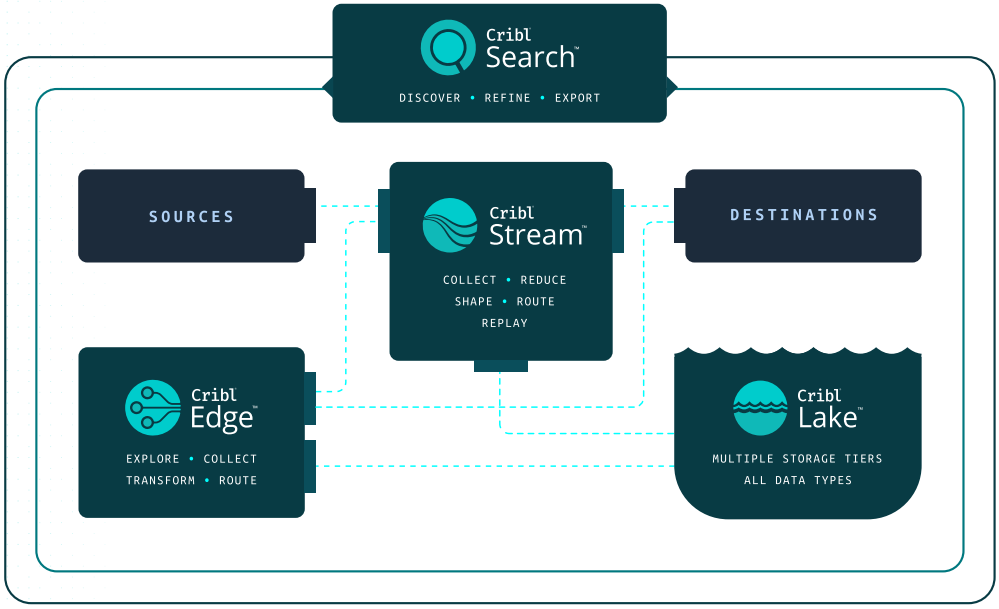


Fig. 01:
High-level solution architecture.

Solution benefits.

- **Flexibility to eliminate tool lock-in:** Choose the best vendor tool(s) without requiring new agents or collectors.
- **Simplify observability data engineering**: Easiest way to discover, explore, collect, enrich, and route data from anywhere to anywhere.
- **Complete control over your data:** Route data where it has the most value. Control access to sensitive data. From core to the edge.
- **Observe more:** Build and automate tiered storage. Search data in place, only forward useful data tailored to your needs. Access data previously inaccessible.
- **Stretch your investment:** Reduce management costs and complexity of exploring, collecting, processing, and accessing data at scale.

Product overview.

Cribl Stream.

Functionality.
Cribl Stream is an observability pipeline that gives you the flexibility to route, shape, restructure, and enrich data from any source to any destination without adding new agents. Gain control over your data and simplify your IT and security efforts. Instrument everything, analyze more data, and pay less.

Customer needs addressed.
Stream empowers organizations to take control of their data in motion. It's a powerhouse for routing, reducing, and enriching data, making it more manageable and insightful. By streamlining data flows, Cribl Stream ensures that your data is not just voluminous but valuable, cutting through the noise to deliver clarity and actionable insights.

Cribl Edge.

Functionality.
Cribl Edge is a highly-scalable edge-based data collection system for logs, metrics, and application data that enables administrators to reliably collect and process logs, metrics, and application data in real time from Windows or Linux machines, apps, and microservices and deliver them to any supported destination.

Customer needs addressed.
With Edge, Cribl brings processing power closer to where data originates. This decentralized approach to data collection minimizes latency and maximizes efficiency, ensuring that data is optimized before it travels across your network. It's about making smarter use of resources and ensuring that your data infrastructure scales intelligently with your needs — all the while providing a centralized management view for insights and efficiency.

Cribl Search.

Functionality.
Cribl Search is a vendor-agnostic analytics tool that enables teams to perform 'search-in-place' queries for more rapid and cost-effective data utilization, whether that data is at the edge, in flight, in an observability lake, or within existing systems.

Customer needs addressed.
Revolutionizing data exploration, Cribl Search offers a unique lens through which to view both historical and real-time data. It bypasses traditional barriers, allowing for swift, on-the-spot insights without the need to relocate or duplicate data first. This agility transforms how decisions are made, grounding them in the most current and comprehensive data view.

Cribl Lake.

Functionality.

Cribl Lake serves as a central hub for storing, managing, and accessing vast amounts of data collected from disparate sources. It's designed to complement the Cribl ecosystem, seamlessly integrating with Cribl Stream and Edge to ingest and route data downstream. With Lake and Search, you can query data at rest, in and across Cribl Lake and all data lakes, object stores, search APIs, and analytics solutions. Federated search allows you to query not just one set of data but multiple data stores and sources simultaneously. This means you can run fast investigations and get value from data without delays.

Customer needs addressed.

Lake addresses the critical need for organizations to not only collect and analyze data without having to move or rehydrate data,but also to retain valuable information for future use. It provides a reliable foundation for historical data analysis, compliance, and security investigations, eliminating the complexity and cost typically associated with large-scale data storage solutions. By offering a unified platform for data from ingestion to long-term storage, Cribl Lake ensures that valuable insights are never lost and always within reach, empowering organizations to make informed decisions based on comprehensive data landscapes.

The below diagrams give closer looks at the processing and transformation options that Cribl Stream and Cribl Edge provide internally. Sources collect data, and get it to Destinations via QuickConnect or Routes, which manage data flowing through Pipelines. From there, users can easily store any data in Lake. Additionally, users can transform and analyze log events in any destination with Cribl Search.

NOTE: Cribl Stream may serve as an intermediary destination for Cribl Edge.

Fig. 02:
Cribl Stream basic solution architecture.

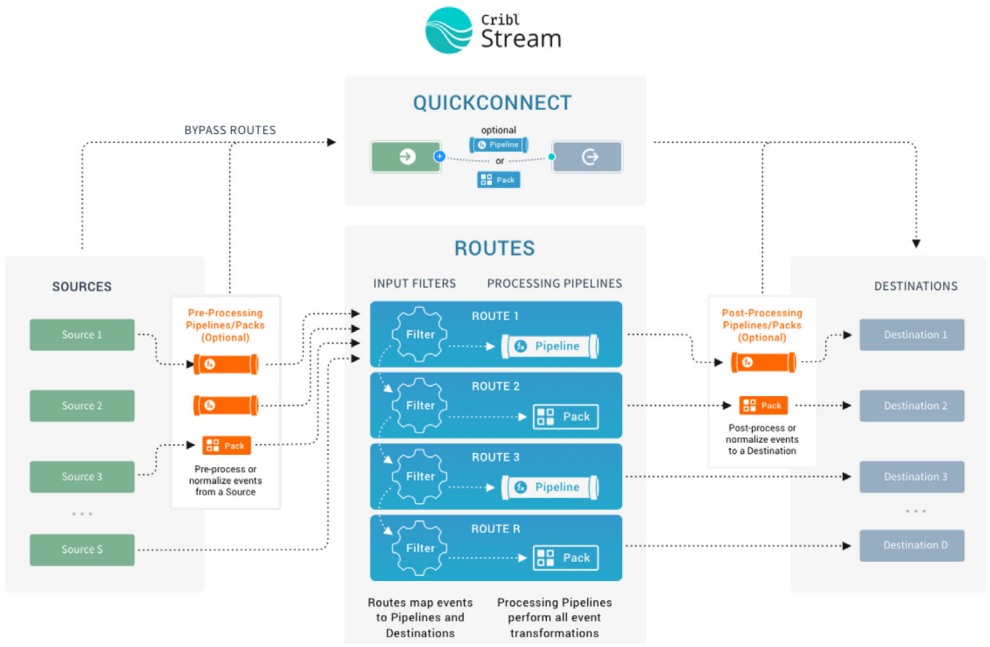


Fig. 03:
Closer look at a single Stream Worker (instance).

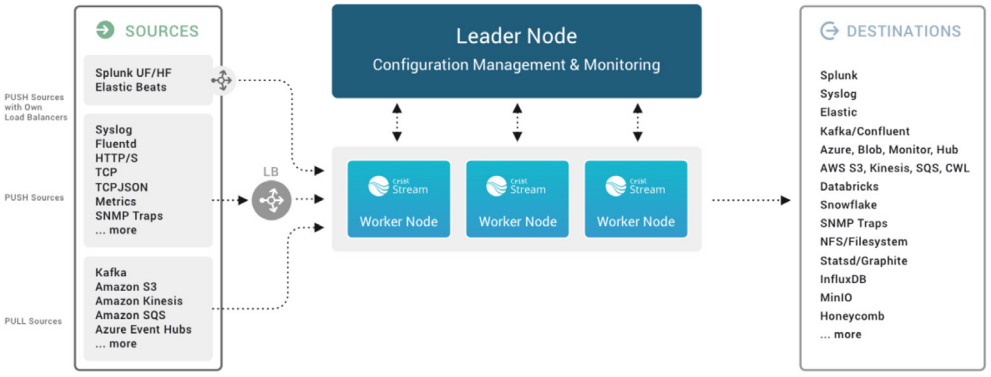


Fig. 04:
Cribl Stream scales up to meet enterprise needs in a distributed deployment.

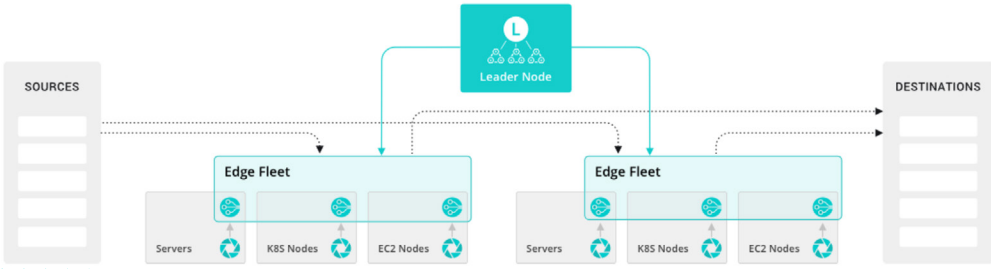


Fig. 05:
Cribl Edge basic solution architecture. Edge processing, management, and receivers.

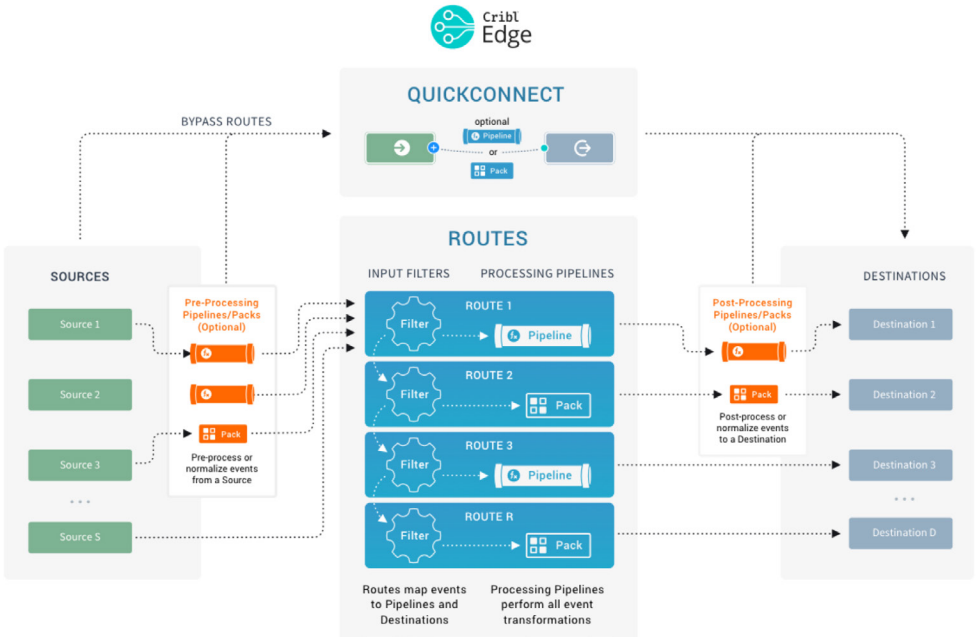


Fig. 06:
Double-click into a single Edge node.

Key capabilities:

Explore data first, then collect.

- Cribl's solutions redefine data management and offer a fresh approach to observability, starting with a unique 'Explore First, Collect Second' approach. This allows for in-depth examination of data at its source, ensuring only relevant data is processed, enhancing efficiency and reducing costs. The integration of Edge and Stream technologies facilitates seamless data handling, from exploration to collection, supporting diverse sources and scales.
- The Cribl solution leverages the power of Edge and Stream to explore logs, metrics, and application data at their egress points, giving you the option to dive into the data before deciding to collect and process it. Collect logs, metrics, and application data from any source at unprecedented scale — and even schedule batch collection from multiple APIs. Use ad-hoc data collection to recall data from your lake or storage solution, and replay it to your analytics tools.

Cost-effective routing and processing wherever you need it:

- Cribl's architecture ensures cost-effective data routing and processing across your infrastructure, optimizing resource use.
- Cribl Edge enables data reduction, transformation, and routing at the edge, providing more flexibility and lowering costs for forwarding and storing data. Forward data to Cribl Stream for additional processing or land it in the destination of your choice.

Moar data value through long-term data retention and analysis:

- Cribl Lake adds a crucial function to our suite by offering a dedicated space for keeping large amounts of data secure and accessible over time. This is especially important for businesses that rely on historical data analysis to inform future decisions, comply with regulatory requirements, and conduct thorough security investigations.
- Lake enhances the data lifecycle management by providing a scalable and secure environment for data at rest. Lake works smoothly with the data processing of Stream and the data collection of Edge, creating a full-circle approach to handling data from the moment it's created to when it's stored for the long term. This integration allows organizations to not only explore and collect data efficiently but also to retain and analyze it effectively for a complete understanding of their data landscape.
- Lake extends the life and value of your data and simplifies the complexities associated with large-scale data storage and retrieval. Whether it's performing time-series analysis, conducting retrospective audits, or powering advanced ML models, Cribl Lake ensures that your data is always ready and accessible, empowering your organization with historical context and deeper insights.

A "Search Then Forward" approach:

- With 'Search Then Forward,' Cribl Search introduces an innovative way to interact with data across various storage solutions, enabling precise, risk-free queries before data collection. Search's query capabilities can federate searches across multiple data types and multiple data stores, prior to collecting and storing the data.
- Eliminate risk and uncertainty by querying data wherever it lives – at the edge, moving through a pipeline, stored in a data lake, or kept in TSDBs or log stores. With this search-in-place approach, Cribl introduces an innovative way to interact with data across various storage solutions, enabling precise, risk-free queries before data collection.

Overall, Cribl provides more data choice and flexibility. When organizations use Cribl to route data where it has the most value, complex projects like scaling with multiple tools or infrastructure consolidation become a lot easier. Cribl's end-to-end observability solution enables teams to easily control access to or sanitize sensitive data in flight, reduce security risk, and support compliance efforts, giving users better data visibility and reliable analytics while making the most of budgets for licensing, hardware, storage, and people.

Solution deployment.

Before you deploy: Key considerations.

Before you get up and running with the Cribl suite of products, you need to determine the type of deployment that would best fit in your environment. Here are a few things to consider that will help you in your deployment planning:

- **Amount of data ingest:** This is defined as the amount of data planned to be ingested per unit of time. How many MB, GB, or TB / day?
- A**mount of data to be collected at the endpoint:** What amount of that data is planned to be ingested at the endpoint per unit time?
- **Amount of data processing:** This is defined as the amount of processing that will happen on incoming data. Are there a lot of transformations, regex extractions, parsing functions, field obfuscations, etc.?
- **Routing and / or cloning:** Is most data going to a single destination, or is it being cloned and routed to multiple places?
- **Deploying onto servers with no internet access:** Do you need to accommodate air-gapped on-prem servers (i.e., servers with no internet access)?

Once you understand these key points and answer the questions above, it'll be easier to figure out which deployment is the best fit for your use case.

Deploying the Cribl Suite of Products.

This part of the guide focuses on how to set up the Cribl suite of products—Stream, Edge, Search, and Lake— in a way that fits best with what your org needs. We'll look at different deployment models, from single-instance setups to distributed environments, and highlight just how easy it is to integrate Lake for long-term data storage and accessibility.

Types of deployment:

- Use Single-Instance Deployment when incoming data volume is low, and/or amount of processing is light. For guidance, check out our Getting Started Guide.
- Use Distributed Deployment to accommodate increased load. (See Sizing and Scaling for detailed guidance. See Bootstrap Workers from Leader to streamline Workers' deployment via scripting.) For help, take a look at our Distributed Quick Start.
- Use Cribl.Cloud to quickly launch a Cribl-hosted deployment of the combined Cribl solution (Stream, Edge, and Search). With this option, Cribl assumes responsibility for provisioning and managing all infrastructure, on your behalf.

Setting up Cribl Lake for long-term data management

Cribl Lake is built to store and manage large volumes of IT and security data to be retained for the long haul. Integrating Lake into your environment involves some key steps to get your data flowing and stored seamlessly:

1. **Direct Data to Cribl Lake:** Configure Cribl Stream or Edge to route necessary data streams to Cribl Lake for storage in open formats. This configuration is pivotal for capturing the right data for your long-term needs.

2. **Activate Data Collection from Cribl Systems:** Cribl Lake can automatically collect essential operational data from your Cribl deployments. This capability is essential for a comprehensive view of your system's health and performance.

   To activate this feature:
   - Visit your Cribl.Cloud Organization's homepage.
   - In the Cribl Lake section, enable the option for internal data storage.

3. **Organize Data with Datasets:** Cribl Lake uses datasets to organize and manage stored data, improving both efficiency and accessibility. Start with the default datasets and consider creating custom ones tailored to your specific requirements.

   To create a custom dataset:
   - In the Cribl Lake interface, choose to add a new dataset.
   - Provide a distinct name and a description for your dataset.
   - Save your changes to make this dataset ready for use.

4. **Replaying Data via Cribl Lake Collector:** For scenarios where you need to access and reprocess stored data, set up the Cribl Lake Collector in Cribl Stream. This setup enhances the value of stored data, making it more than just an archive.

5. **Choosing Cribl.Cloud for Deployment:** Cribl Lake is designed for the cloud and is an integral part of Cribl.Cloud. This cloud-based approach simplifies your deployment, offering a scalable and secure environment without the complexity of traditional setups.

Incorporating Cribl Lake into your deployment enhances your data strategy by ensuring valuable data is not only captured and stored but remains accessible and actionable for future needs.

Recommended solution deployment option: Cribl.Cloud.

The Cribl.Cloud platform quickly spins up all Cribl products — Stream, Edge, Search, and Lake — in just a few minutes. This deployment option puts the Leader and the Worker Node in Cribl.Cloud, where Cribl assumes responsibility for managing the infrastructure, simplifying deployment and adding flexibility. (If needed in the future, you can expand this to a hybrid deployment of your choice with any desired complexity.) To get started, check out our Launch Guide and sign up on the Cribl.Cloud portal.

Summary.

Let's face it: Running an effective data engine is starting to seem impossible. Data volumes are growing astronomically, leading to higher licensing expenses, rising infrastructure costs, and data retention.

Business needs are changing too, and they're getting more complex. Not only does security continue to be a concern, but companies also need to make sure they can move with the market while maintaining the data visibility they need for effective analysis. Proprietary tool sprawl can make it hard to get data from one place to another or use data sets across teams and platforms.

Transforming IT and Security practices with Cribl's data engine addresses rising data challenges and complex business needs. When you simplify data flow with a processing pipeline and introduce top-tier search capabilities into your approach, you regain choice and control over all your data.

Cribl's end-to-end observability solution combines data collection, routing, processing, storage and analysis for ultimate data visibility and control – while reducing costs and complexity. Deployable in the cloud, on-prem, or using a hybrid approach, the Cribl solution:

- Leverages the power of Edge and Stream to explore logs, metrics, and application data at their egress points, giving you the option to dive into the data before deciding to collect and process it.
- Enables data reduction, transformation, and routing at the edge, providing more flexibility and lowering costs for forwarding and storing data. Forward data to Cribl Stream for additional processing or land it in the destination of your choice.
- Makes it easier to manage your data from start to finish with Cribl Lake, keeping it safe and easy to access for long-term retention. By working hand-in-hand with Cribl Edge and Stream, Lake helps your team not only gather and explore data but also keep it for thorough analysis later on, ensuring you have all the information you need to make smart decisions and stay compliant.
- Eliminates risk and uncertainty with search-in-place query capabilities that can federate searches across multiple data types and multiple data stores, prior to collecting and storing the data. Effortlessly search your Cribl Lake or sift through data in other major object stores.

The fastest way to get started with Cribl is in the cloud! The Cribl.Cloud platform quickly spins up the Cribl solution – including Stream, Edge, and Search – in just a few minutes. To get started, check out our Launch Guide and sign up on the Cribl.Cloud portal.

Additional Reources:

- Cribl Documentation
- Getting Started Guide
- Distributed Quick Start
- Launch Guide
- Cribl Community
- Cribl Support
- Cribl University
- Cribl Curious
- Self-Guided Trials
- Cribl Github Repos
- Docker Hub

ABOUT CRIBL

Cribl, the Data Engine for IT and Security, empowers organizations to transform their data strategy. Customers use Cribl's vendor-agnostic solutions to analyze, collect, process, and route all IT and security data from any source or in any destination, delivering the choice, control, and flexibility required to adapt to their ever-changing needs. Cribl's product suite, which is used by Fortune 1000 companies globally, is purpose-built for IT and Security, including Cribl Stream, the industry's leading observability pipeline, Cribl Edge, an intelligent vendor-neutral agent, Cribl Search, the industry's first search-in-place solution, and Cribl Lake, a turnkey data lake. Founded in 2018, Cribl is a remote-first workforce with an office in San Francisco, CA.

Learn more: www.cribl.io | Try now: Cribl sandboxes | Join us: Slack community | Follow us: LinkedIn and Twitter

©2024 Cribl, Inc. All Rights Reserved. 'Cribl' and the Cribl Flow Mark are trademarks of Cribl, Inc. in the United States and/or other countries. All third-party trademarks are the property of their respective owners.

SDGE-0001-EN-4-0824