Cribl | iHerb

>SOLUTION BRIEF_

From Open Source to Optimized: iHerb's Data Journey with Cribl

iHerb.com is the premier online shop for health and wellness products, offering 30k+ items to 9.5 million customers in 185+ countries. As they've grown, iHerb could no longer dedicate the required time and resources to sustain their open-source data platform. Switching to Cribl has significantly cut costs associated with infrastructure, latency, engineering time, and downtime.

A few years ago, iHerb set out to build a real-time stream processing system for their logging data. However, developing an in-house data pipeline consumed a lot of engineering resources and left them with a lot of technical debt, making the solution costly and unmanageable for the long term. Learn how iHerb replaced a homegrown data pipeline with Cribl Stream.

As an online retailer, iHerb was processing 2–3TB of weblogs daily, so configuring sources and keeping systems up to date was eating up valuable engineering time. Bob Chen, the organization's Senior Director of Infrastructure Engineering, mentioned the other factors that led to the switch from building their own tool to using Cribl Stream:

> **"We wanted an easy-to-use tool without having to tap into a UX team. A good API interface was critical, as was support for multiple logging sources and destinations. It turned out Cribl Stream could provide all that, and it was easy to implement and deploy, so we made the choice to put our build on hold."**
>
> **— Bob Chen**, Senior Director of Infrastructure Engineering

HIGHLIGHTS

- Seamless management of 5TB/day of data with selective routing to appropriate destinations.
- Reduced load on Spunk, Elastic, and Loki via intelligent volume reduction and optimized data.
- Increased uptime by allowing engineers to focus on tasks more critical than building and maintaining tools.

Using Selective Routing to Manage Doubling Data Volumes

The decision to switch from open source to Cribl Stream came at just the right time, as the amount of data iHerb processed daily, doubled. Their data now flows seamlessly from sources like Kafka and Fluentd to destinations like S3, Loki, Elastic Stack (Elasticsearch, Logstash, Kibana), and Splunk.

All that data goes to S3 for long-term storage, with most logs going to Elastic for short-term (<3 months) storage. Some selected logs get sent to Loki for retention periods between 3–6 months. iHerb's Security department provides guidelines to Bob and his team regarding which data gets sent to Splunk for security use cases.

> **"We process a lot of data each day, and we can't afford to skip even a few KB of it — we need every log entry to troubleshoot incidents and identify other issues. Using Cribl Stream helps us avoid losing any of the critical data we need."**
>
> **— Bob Chen**, Senior Director of Infrastructure Engineering

> **"Ultimately, we also ended up as a self-service model for some of our dev teams. They can do a lot on their own, so now I don't get a JIRA ticket to set up any new sources or destinations."**

Reduced Load on Analysis Tools

Cribl Stream's ability to help iHerb be surgical about their data–selective dropping, sampling, suppression of whole events, and routing data to the best tool for analysis–makes Splunk, Elastic, and Loki return searches faster while reducing required infrastructure and processing power. The ability to simply configure pipelines allows for easy reformatting, removal of redundant or unnecessary fields, stripping out null JSON values, and more.

Stream also offers native data transformation functionality, simplifying data management and reducing storage of surplus data. Annotations in Kubernetes metadata can vary in size, and the data is often unstructured. iHerb uses Stream to trim out unnecessary fields and clean up, redact or transform these events.

Improved Security With Masking and Replay Features

With the increase in cybersecurity incidents in recent years, securing sensitive data is more important than ever. iHerb leverages Cribl Stream to mask sensitive patterns using redaction, hashing, or randomization. These functions allow Bob and his team to mask PII for the security team.

If a security incident does occur, Cribl Stream's Replay feature allows them to selectively re-ingest data from S3 back into their systems of analysis. And going forward they'll be able to use Cribl Search, which allows you to search data in place (ie before ingesting into analytics tools), to find investigation-related context from across various S3 buckets.

Observability, Metrics, and Beyond

Many teams leverage Elastic for log analysis, but it's also a popular choice for handling metrics. iHerb uses Cribl Stream to query and aggregate log counts and other statistics based on parameters like cluster, namespace, and source. The results are then routed from Elastic into an intuitive, user-friendly Grafana dashboard, enabling them to gain valuable insights into system performance, identify trends, and troubleshoot issues effectively.

Since successfully implementing Cribl Stream, Bob and team have also used Cribl Edge to implement a couple thousand edge nodes. Cribl Edge is a centrally managed, edge-based data collection system.

Kubernetes, an integral part of iHerb's infrastructure, is notoriously difficult to monitor and often limited by the observability of the system. iHerb deploys Kubernetes with Edge already bootstrapped to collect application logs and system metrics, giving them visibility into Kubernetes microservices.

> "Outages are costly for us as an e-commerce organization — Cribl Stream allows our engineers to use their time and expertise on minimizing downtime and other important tasks."

> **"The combination of Cribl Stream and Edge is a lifesaver. The speed, accuracy, and ability to manipulate logs is unparalleled."**
>
> **— Aaron Wilson**, Senior Site Reliability Engineer

By using Cribl Stream and Edge instead of building their own infrastructure, iHerb has been able to save on infrastructure, network bandwidth, engineering, and outage costs — and the setup was even easier than Bob and his team anticipated.

> **"We got our Cribl Stream POC up and running within a week. We tested as many scenarios as we could, pushed a bunch of our logs through a test environment, then made the purchase and got our production environment going remarkably quickly."**
>
> **— Bob Chen**, Senior Director of Infrastructure Engineering

> "Cribl Stream paid for itself, Cribl Edge pays for itself."

TL;DR

- iHerb switched from building their own observability infrastructure to using Cribl Stream
- Reduced volume of data stored via selective dropping, sampling, and suppression of whole events.
- Secured sensitive data using redaction, hashing, or randomization with Cribl Stream's Mask function.
- Reduced load on Splunk, Elastic, and Loki by intelligently filtering which logs end up in each destination.
- Aggregated logs and metrics to create user-friendly dashboards in Grafana.
- Leveraging Cribl's portfolio of products — Cribl Stream, Search, and Edge — gives them control of what data they collect, and helps to save time and money by routing data to the best, most cost effective place for storage, analysis and compliance.

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.

CS-0004-EN-2-0624