Cribl | SOPHOS

>CASE STUDY_

How Sophos modernizes data management with Cribl

HIGHLIGHTS

- Enabled data independence, allowing flexible ingestion of diverse custom log formats from Sophos products
- Enhanced data cataloging and provenance tracking, improving visibility and context surrounding data assets
- Exceeded target to reduce costs by 30% (hit 48%!) through expedited data onboarding and optimization

Sophos

Sophos protects more than 600,000 organizations and 100 million users across the globe with its wide array of cybersecurity products and services. But the company's internal security team must ingest data from a wide array of sources to power their defenses to ensure Sophos is safe from cyberthreats.

Chris O'Brien is VP of Security Operations for Sophos, a global leader providing cybersecurity products and managed services. Chris and his team's job is to defend Sophos by identifying, mitigating and responding to cybersecurity incidents against the company and its products and services. A big part of doing that is sifting through terabytes of log and security telemetry from myriad sources daily to find malicious needles in the data haystack. And Cribl has allowed the team to efficiently process data through several different pipelines, making their work more cost-effective and allowing their analysts to spend their time where they can have the largest impact.

When he joined the team, Chris found the team was already exploring Cribl Stream as a potential solution to their big data challenges. Initially skeptical, Chris asked his team to justify their decision. However, after seeing the proof-of-value (POV), he was convinced.

> "The team convinced me by showing me the POV, how the data was being funneled through various filters, how we were processing it, routing it, and how we got the clean data we needed to drive our investigative processes. It was immediately clear that Cribl was saving us money and space and was generally doing a good job"
>
> —**Chris O'Brien**, VP of Security Operations at Sophos

Significant reduction in data volume

Sophos wanted to reduce the volume of data it was processing and storing, as the company relies on many data sources to refine its threat intelligence and make Sophos products more secure. One data source was a particular challenge, contributing nearly a quarter (1TB) of Sophos' daily data intake.

The Security Operations team knew that much of the data they pulled back from this product was unnecessary but did so anyway because their existing infrastructure could not filter it at the point of ingestion. Instead, they had to ingest all the data, then manipulate and process it downstream in their data lake. However, by then, they had already incurred infrastructure costs.

Cribl Stream data processing capabilities allowed Sophos to quickly identify and filter out unnecessary fields, resulting in a 25% reduction in data volume from that one product.

> **"Our Technology Lead implemented the filter he had prepared in the POV, and our data ingest for that product fell by 25% in the first few weeks. We aimed to achieve that goal by the end of the quarter - he'd done it within the first month. We knew it was coming because we saw it in the POV, but it was still quite amazing to see it actually happen on paper."**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos

Based on the POV's performance, Sophos set a KPI of reducing overall data volume by 30% by the end of the financial year. They exceeded this goal in a matter of months, ultimately bringing down overall data volume by nearly 50%.

Flexible ingestion of diverse log formats

However, Sophos quickly realized they could go beyond just data volume reduction; with Cribl Stream, they could take on data normalization, too. Each of the data sources that Sophos ingests, and often even different versions of the same data sources, has its own logging format, making it difficult for the Security Operations team to normalize and process the data effectively.

To overcome this problem, the company explored traditional data normalization tools. However, these tools created further problems, forcing Sophos to rehash logs to conform to a standard format. This process would require significant engineering effort and internal struggle to convince engineers to change their logging practices.

> **"Beyond the standard logging that you'd expect of a corporate environment, we also handle a ton of custom logging formats at Sophos. That's great, because we can get out of the way of the engineering teams and do the mapping in the back end. But to do that requires proper visibility of the upstream pipeline and the ability to tweak and change things on the fly. That's not common in most security platforms without having to roll deep on DevOps"**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos

Cribl offered Sophos the necessary flexibility to ingest these diverse data logs without having to enforce strict data standards on their engineering teams.

> **"Cribl gives us the independence to allow people to generate logs how they want to generate them. Then, from a security perspective, we can coordinate those logs, filter them, parse them, change them to fit specific security use cases, and finally send the processed data to where it needs to be–all from Cribl's clean user interface."**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos.

In essence, Cribl provided Sophos with the data processing choice and flexibility they needed to take raw log data from various sources, filter and normalize it, and then route it in a usable form to support their security operations and analytics.

Data cataloging and provenance tracking

Before using Cribl, Sophos' need to ingest and analyze masses of data outpaced its ability to implement a more structured data infrastructure. As a result, it ended up with an "accidental data lake" created by aggregating data from various sources over time without a clear architectural plan. This infrastructure resulted in a lack of visibility into the provenance and lineage of the data they were collecting, making it difficult to understand and manage.

> **"We were just ingesting all the data and then aggregating it into a big data store. This approach works, but after a while, we began to ask questions like: Where did that data come from originally? What was its origin? How did it get merged with this? Before implementing Cribl, we couldn't answer those questions."**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos.

This lack of visibility led to a concern that team members could compromise the integrity of sensitive customer data by merging it with other data sources. As Sophos started using Cribl to ingest and process data, it automatically created a data catalog and provenance chains.

> **"When you start plugging data feeds into an upstream data preprocessor, you automatically create a data catalog and provenance chains for all those feeds because you're building the pipelines, you're building all the steps that that data is going through. That might sound obvious, but it just wasn't something we had with our previous data lake setup."**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos.

Planning for future improvements

While it focuses on getting all its data through Cribl, Sophos has big plans to modernize its data strategy. The data flexibility and control provided by Cribl have offered new opportunities to evolve Sophos' data and security infrastructure beyond its "accidental data lake" setup.

> **"When pre-processing data with Cribl, you have opportunities for enrichment, correlation, and pre-processing in-stream. Then, you can do it again when the data is at rest and again when you get other data coming in. It starts to open this multi-dimensional tech stack where you can plug and play different technologies and then do cool things with it."**
>
> **—Chris O'Brien**, VP of Security Operations at Sophos.

To discover how Cribl can modernize data management for your organization, including by reducing data volume, improving data normalization processes, cataloging data, and more, schedule a custom demo today.

TL;DR

- **Achieved a 48% reduction in overall data intake, exceeding the 30% target in the first 30 days.**
- **Sophos struggled with the massive scale and variety of security and logging data.**
- **Faced challenges around data normalization, processing, and analytics.**
- **Enabled flexible ingestion of diverse custom log formats.**
- **Automated data cataloging and provenance tracking for improved visibility.**
- **Sees Cribl as foundational for evolving beyond "accidental data lake."**

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

©2025 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-0037-EN-1-0325