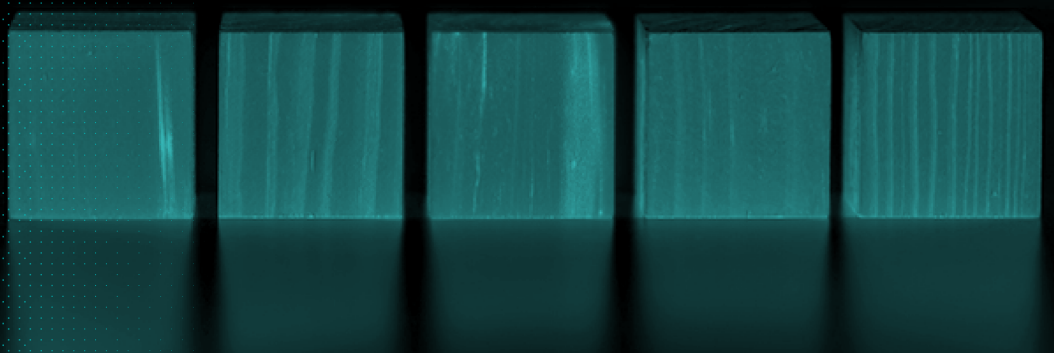
Cribl

>WHITE PAPER_

Five ways to get even more out of Elasticsearch.



>WHITE PAPER_

Five ways to get even more out of Elasticsearch.

Introduction.

In today's digital world, businesses face a big challenge: data is growing incredibly fast, much faster than IT and security budgets can keep up with. Imagine trying to fill a growing number of water balloons with the same small hose; eventually, you won't be able to keep up.

This is where smart data management comes into play. It's like finding more efficient ways to use the hose you have or finding better balloons that need less water. We're talking about making the most of every piece of data, using smarter storage solutions that don't break the bank, and ensuring our tech can handle today's needs and tomorrow's growth. Teams need a smarter, more flexible approach to handling the massive waves of data, ensuring they're ready for the future without spending a fortune.

How can you take advantage of the highly performant queries of Elasticsearch, support observability efforts, and reduce your budget?

While indexed logging tools like Elasticsearch are optimized to perform very fast searches, they can be inefficient for long-term data retention, requiring organizations that use them to store large data footprints. This contributes to ballooning infrastructure costs as these larger data stores lead to accelerating storage costs and require more CPU to power these searches. How can you take advantage of the highly performant queries of Elasticsearch, increase the depth and breadth of your observability efforts, AND reduce your budget?

With help from Cribl, you can trim storage and compute costs dramatically and get the most out of Elasticsearch. Simply follow these five steps:

1. Leverage existing data sources and agents to route data to Elastic.
2. Achieve big savings by retaining IT and security data in lower-cost object storage.
3. Make it easier to migrate your data from older versions of Elastic.
4. Eliminate data with little analytical value, to control costs.
5. Easily navigate data migrations at scale from one tool to another.

Seem like a tall order? Read on to find out how to get started.

Leverage existing data sources and agents to route diverse data to Elastic.

Think of your data as a gold mine that's already full of nuggets; you just need to find and use them smartly. When you route or receive data with Cribl's vendor-agnostic solutions, you can rest assured you're not letting any of that gold go to waste. The key is using every bit of data you've already got in new and useful ways. Sometimes, this means getting your data to the right tools to dig through that mine, find the gold hiding in the corners, and make sure it's put to good use. Adding new tools and destinations to your observability environment gives you the flexibility to choose the best tool for the job. In the past, this usually meant deploying agents and collectors to route data to each tool in its unique, proprietary format. And for organizations with thousands of endpoints, deploying agents for a new tool can take months.

An easier way is to reuse the data that you already have by shaping it into the formats required by any new tool. Cribl allows you to route data to any tool without having to deploy new agents or re-train users. It's not just about saving resources by reusing what we have; it's also about making smarter decisions because we're using all the valuable information available to us. This way, you can finally see and access the full value of your data, ensuring nothing valuable is overlooked.

Retain a full-fidelity data copy in a lower-cost cloud storage, but still be able to replay and analyze it at will.

Achieve big savings by retaining data in lower-cost object storage.

Data storage costs can quickly spiral out of control, like a shopping spree without a budget. Most organizations keep several months of data in their logging systems to fulfill compliance requirements, or in case they need to analyze it at a later date. Keeping large amounts of data in logging systems is very expensive and hurts query performance. Adding compute power to combat this performance degradation has diminishing returns and escalating costs. The increased costs and bloated infrastructure outweigh the convenience of having this data on hand and ready to analyze.

There are, however, many good reasons to make data available for relatively longer retention periods. Analyzing trends, conducting investigations, and meeting stringent compliance requirements are just a few. The systems you use to analyze data are not the best places to store it. The vast majority of queries in logging systems are on data from the past few days. Queries on older data happen, but they are typically less time-sensitive. They don't require that you store data in block storage, which can cost as much as 100 times more than cloud and other object storage options. Object storage options cost significantly less than storing data in your logging infrastructure. These options include file systems, dedicated data lakes, and cloud solutions like AWS S3, MinIO, and Microsoft Azure Blob Storage. Choosing these locations for longer-term storage of IT and security data can equate to a fraction of the cost of keeping them in indexed logging tools.

Enable administrators to easily provision datasets to their security and operation teams.

Cribl works with both, unlocking smart ways for teams to store data without emptying their pockets. Now, you can choose the right mix of top-dollar, high-speed capabilities for the data you need every day, and more affordable, object storage for data you don't need as often.

From there, Cribl's Replay feature lets you use object storage as a simple, efficient, and low-cost storage system to retain data that you aren't analyzing today, but may need later. Using Replay, you can collect data from object storage, filter, enrich, right-size, and format it in flight for Elasticsearch (or for any other analytics tool you need to use) whenever you want it. TransUnion saved $500K on Splunk licensing costs in 30 minutes by using Cribl Stream to send a full-fidelity set of their data to lower-cost, longer-term storage for potential future review, and then filtering out and stripping out unnecessary data.

Cribl makes it possible for your data, tooling, and team to meet both your immediate business requirements and your long-term needs, making sure you can keep and use all your important data without spending a fortune.

Make it easier to migrate your data from older versions of Elastic.

One of the biggest challenges of maintaining a long-term analytics program using Elastic has been migrating to newer versions of Elastic. It's kind of like renovating a house while living in it: The end result is super appealing, but the process of getting there can make day-to-day living difficult. Put more explicitly, new versions bring powerful new capabilities, but they also complicate how data needs to be collected and formatted to even take advantage of these new features.

Cribl makes it possible for you to keep and use all your important data without spending a fortune.

The Elastic Common Schema (ECS) has changed several times over the past few years. Every time the schema changes, the Elastic tooling requires data to be formatted to meet the schema changes. ECS can change independently from tool versions, and the agents you have installed today most certainly will not collect data in the ways ECS updates will require in the future.

Cribl solves the problems presented by the evolution of ECS. Rather than constantly rework how you collect data to keep pace with ever-changing requirements and newer versions of Elastic products, Cribl shapes and transforms data into the required formats in real time as it streams. There is never a need to deploy new agents and collectors, or to tweak existing ones, in order to analyze the data your organization needs. Cribl receives data in whatever shape it exists today and helps you transform it into the shape, size, and format that you require. This future-proofs the investments you've made in Elastic by ensuring that the next time ECS changes, you can easily shape your data to fit.

Eliminate data with little analytical value to control costs.

Not all data is created equal; some of it's just noise that drowns out the valuable insights. How can you filter out that noise, and keep only the data that truly matters? Cribl can help declutter your data, that is, remove fields, values, and datasets you no longer need to make room for what's valuable. By eliminating low-value data, businesses can focus on the information that drives decisions, improves services, and enhances customer experiences.

Ops professionals typically use only a fraction of the data they ingest into tools like Elasticsearch. And from what we've seen with our customers, as much as 50% of log data has little or no analytical value (when was the last time you wanted to analyze null values and duplicate data and fields?).

Because you can keep a copy of this log data in object storage, you don't need to worry about missing something by being overly aggressive in reducing your log volume. And, you can always replay it to Elasticsearch or another tool if you need it later.

Just ask Autodesk's data analytics engineering team: They used Cribl Stream to reduce the ingestion of duplicate data across all of their logging and SIEM platforms by nearly 15%.

To achieve similar results, here are three techniques for reducing junk data (and not having to pay to store and analyze it) in your IT and Security operations:

Dropping unnecessary fields.

The developer who originally builds a log message has a strong incentive to stuff that message with as much information as possible to avoid having to go back and add more information later. However, for consumers of this data, especially from devices and software outside of the organization's control, much of this information is unnecessary. In Cisco eStreamer, for example, every message contains dozens of key-value pairs that are set to NULL or N/A. Microsoft PowerShell logs contain a full copy of the script in every log message. Preprocessing this data to remove unnecessary fields from these messages can result in dramatic savings.

Converting logs to metrics.

Many of the highest-volume data sources are really metrics in disguise: web activity logs, network flow logs, or custom application telemetry. If a measurement value in the log is the reason for ingesting it, then it might make sense to aggregate those logs into summary metrics using Stream. Metrics can still be stored in a logging tool, often with dramatic reduction in event counts and data volume. Metrics can also be sent to a dedicated time-series database for efficient storage and retrieval.

Deduplicating event streams.

Log data often contains a lot of repeat information. This can include duplicated fields (like multiple timestamps conveying the exact same information to your analytics environment) or entire log messages that emit repeated information. Duplicate data cannot tell you anything new about your environment, so Stream helps you drop the duplicates for dramatically lower data ingestion.

> "Lots of data was being duplicated in our logging platforms, and Stream allows us to easily detect when this happens. We were able to reduce extra ingestion by 93.1 percent – an internal target we set for ourselves by just doing the easy stuff!"
>
> — Sudha Kanupuru, Cloud Architect
>
> AUTODESK

Easily navigate data migrations at scale from one tool to another.

Often, a new team at an organization wants access to observability and or security data for a new project. So, they turn to the free versions of Elasticsearch, Logstash, and Kibana to see how well the Elastic Stack performs against their goals before committing to the commercial versions of the products. But, the allure of flexibility and ease often overshadows the complexity of adding a new tool, especially for larger projects where the scale of data collected, stored, and analyzed can get overwhelming. (Switching to a new phone always sounds appealing... until you remember all your contacts, photos, and apps need to move too.)

It's not easy to make that switch smooth for your data. It's about having the right tools and processes in place so that when you decide to upgrade or change tooling, the transition is seamless. This means your data is always ready to move where it's needed most, without losing time or information. Like your business, your data should be agile and ready for new opportunities.

Cribl can dramatically simplify the process of migrating from one tool to another — like Splunk to Elastic — or migrating workloads to the cloud. In fact, one of the Big Four used Cribl Stream to simplify the entire company's cloud migration and cut back on the data they needed to send to achieve the same business results (Here's the kicker: Before Stream, they streamed 5-6TB of Windows Event Logs to the cloud; after using Stream, they cut that volume by 20% with the dedupe function alone!).

We mentioned earlier how Stream can reshape data between different versions of Elastic and ECS. Similarly, Stream can take data collected by a Splunk Universal Forwarder and shape and transform that data for the Elastic. Stream can also repurpose logs and metrics collected by Beats or Fluentd agents for other analytics and storage destinations.

Ultimately, all of this gives you the flexibility to use whatever data you have today, and process it for use for all of the tools in your stack as destinations. This flexibility can alleviate many of the concerns organizations have around being locked in to one or a handful of expensive vendors.

Summary

In wrapping up, the white paper emphasizes strategic approaches to leveraging the Elastic Stack more effectively. The document underscores:

1. **Leverage existing data sources and agents to route data to Elastic.**
2. **Achieve big savings by retaining IT and security data in lower-cost object storage.**
3. **Make it easier to migrate your data from older versions of Elastic.**
4. **Eliminate data with little analytical value, to control costs.**
5. **Easily navigate data migrations at scale from one tool to another.**

Elasticsearch is a powerful platform. As with any platform or combination of tools, you need to constantly evolve your data engine to successfully execute your IT and security projects. You may be asked to solve new challenges without a lot of additional budget, requiring you to make sure your tools can keep pace with changes in your business. Adding to this, data volumes are growing, and organizations are being asked to analyze new types of data without adding new staff and infrastructure.

Cribl can help future-proof your data efforts and strategy while greatly enhancing your existing investments in platforms like Elastic. Cribl gives you the flexibility to collect, route, shape, restructure, enrich, and query data at or from any source to any destination, without adding new agents or re-training users.

Cribl gives you the flexibility to choose the tools you and your team need and want. And Stream gives you more control over your data for better outcomes and lower costs — not to mention a simpler, faster way to get any data into the formats needed for any destination. By focusing on reusing data, optimizing storage costs, streamlining migrations, and eliminating non-essential data, businesses can enhance their operational efficiency and analytical capabilities.

Simply put, Stream lets you observe more with Elastic, while paying a lot less and massively reducing the effort of your teams. The result? You and your organization will be ready to adapt to the evolving data landscape with agility and foresight, ensuring you remain competitive and data-driven in your decision-making processes.

To get started with Cribl and Elastic today, **click here to sign up for Cribl.Cloud**. The **Cribl Slack Community** is also a great place to connect with leaders from other organizations leveraging both Cribl and Elastic.

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, and Cribl Search, the industry's first search-in-place solution. 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

WP-0015-EN-1-0224