The Economics of Observability Data

How to correct 3 common log management mistakes.

splunk>

In the modern and evolving technology landscape, businesses face a myriad of challenges that can affect their bottom line.

Legacy monitoring simply isn't up to the task. New applications are now more complex, and it's more difficult to identify and solve problems quickly. All the while, customers continue to demand perfection — and this isn't likely to change anytime soon.

That's why observability matters. The purpose of observability is to help practitioners build resilient systems so businesses can bounce back more quickly when things go wrong. Unplanned downtime can be extremely costly, and that's why software developers, site reliability engineers and IT operations teams must have a unified strategy.

So how can businesses move away from legacy monitoring and toward full-fidelity observability, keep a close eye on all operations, and somehow save time, money and resources? The answer is multifaceted, but to start, it's crucial to understand and properly utilize the pillars of observability: Metrics, events, logs and traces.

In this ebook, we'll explore how each of these building blocks serves its own purpose, the most common mistakes we see practitioners making when it comes to managing logs, and steps your organization can take to achieve more efficient observability operations.

To build our foundation, let's start with this: **Stop using logs for everything. (Use them where they make sense, and keep them stored if you need them.)**

Architecture has evolved over centuries. We use different materials as new information comes to light. We sometimes build log cabins, but we also build commercial real estate. The highest skyscrapers in the world are justifiably built with an elaborate combination of materials, including steel, concrete and stone. And just like a well-built skyscraper that soars hundreds of feet into the air, a strong observability engine requires a strategic plan — and logs shouldn't be the only material holding it together.

Although it's not a best practice, it's a common practice for practitioners to use logs for everything. Because nearly all applications log information about what they are doing, it can be a valuable resource for app developers — but a voluminous one. Oftentimes, the sheer volume of logs exhausts time, energy, money and resources that could (and should) be used elsewhere. Instead, true digital resilience requires a more in-depth understanding of the specific role and purpose of logs within an observability strategy. Of course, logs may very well be the best choice in some situations (like infra monitoring, for example), but using logs exclusively may not give a comprehensive or timely picture of what's going on in your business' environment.

But as we mentioned, this is a common problem — so if your observability practice is relying too heavily on logs, you're not alone and there are ways to fix it to be more economical. Let's deconstruct the three most common mistakes we see with log management and how to correct them.

Mistake #1

Not knowing when to convert a log to a metric

Where logs provide in-depth information about resources, metrics provide a snapshot of resource status.

For example, a metric may show a resource is down, and a deep dive into the logs will help tell you why it went down. While logs can be heavy and can take up a lot of space, metrics are a lighter, more efficient building block of observability. It is essential to know when and why to convert a log to a metric. Using metrics instead of logs in the right places will save size and compute performance while meeting the immediate business need.

A metric, in general, is anything that can be counted, and metrics are most useful for identifying trends over time. This may include infrastructure metrics, error rates or request rates. While you may need to retain full logs in long-term storage for deeper troubleshooting (like for compliance reasons), making this conversion strategically for a real-time view of resource status can be more efficient for your business in the long run.

Since data stored in a metrics format can provide faster search performance and more efficient data storage than you will find with logs, sometimes it makes sense to convert logs to metrics. So how do you know when to convert a log to a metric? Specifically, this makes sense when you have a real-time need and the log can be logically parsed to pull out usable key performance indicator metrics. However, you will lose the full fidelity of the log when reduced to the more usable metric. With the availability of low-cost cloud data lake storage, it's a good idea to retain a full log copy for forensic troubleshooting, compliance or audit. This is helpful just in case the full-fidelity log is needed at a later date with less urgent search performance needs.

PRO TIP

It is essential to know when and why to convert a log to a metric. Using metrics instead of logs in the right places will save size and compute performance.

Mistake #2

Not using traces with your logs

A trace tracks and visualizes the journey of a request as it travels through a complex system of services.

Traces enable you to understand end-to-end transactions aligned to business-level needs (e.g., policy submission transactions) and can be a valuable tool alongside logs and metrics for Application Performance Monitoring (APM) solutions.

Many practitioners who are focused on analytics and retention need an APM solution with the right access to logs, metrics and distributed traces, but instead, they try to generate the concept of a trace from correlation across various log entries. With this approach, you are overlooking the deeper, more flexible insights available through distributed tracing. An ideal solution is a tool that can capture traces. One way to capture and onboard traces into an APM solution efficiently is with OpenTelemetry.

Be sure to choose your provider wisely, as some only provide a sample of the data, and other products become expensive with massive trace volume. It can be challenging to manage and configure this manually, and sometimes you may only be looking at a subset of data.

PRO TIP

Splunk collects, analyzes and stores 100% of traces. Since we don't have access to every transaction, this means you can achieve full-fidelity tracing through Splunk Observability Cloud.

Mistake #3

Not tiering your data

In some cases, it's a best practice to retain all data for as long as possible (like for security or compliance purposes).

However, this does not mean that you should pay for the same level of search performance for every piece of data — especially if your needs are not time sensitive. The volume of your log data will continue to increase over time, but if you're smart about which data matters and when, you can strategically design a storage and search performance trade-off scheme to support your unique needs. This will make it more manageable.

The answer to this dilemma is data tiering, which is the practice of prioritizing and ranking enterprise data from most critical to least frequently used — accounting for geographical, compliance or format restrictions — and categorizing that data appropriately in a cloud data platform or storage system.

There are several undeniable benefits of data tiering. First, it enables you to understand where all of your data lives in a sustainable, manageable way, and also enables you to access the most valuable data faster. With data tiering, you can also pull data with lower, unknown or longer-term value from low-cost stores only when you need it. This optimized access to data leads to improvements in operational efficiency — including more visibility into your environment, improved monitoring efforts and more clarity around your entire data landscape. In addition, data tiering will organize and prioritize data for ingestion based on its value to the business, resulting in more cost-effective data storage.

Of course, data tiering builds resilience because it enables quicker incident investigation and action based on a filtered, transformed or enriched data set in the right formats and at a manageable volume. This means you can move toward a resolution more quickly with fewer resources. Put simply, it helps your business do more with less with just the data you want, in the format you need, when you need it.

PRO TIP

Data tiering is the practice of prioritizing and ranking enterprise data from most critical to least frequently used. It will enable quicker incident investigation and build resilience.

Moving forward

Rather than using logs for everything, it's important to understand when logs provide the most benefit, and when metrics and traces might provide a better understanding of your apps and infrastructure. Data tiering informs which data to keep or store more cost effectively for longer-term use based on ingest costs, search performance, relevance to use cases and other factors. In the future, data tiering will become increasingly important for overall operational efficiency — and tools native to your data ecosystem that allow you to filter, enrich, route and transform data will become a necessity. By incorporating a system that knows how to manage, store and search logs, metrics and traces efficiently, your business will see improved ROI and generate more long-term value.

At Splunk, we believe that data tiering is the best way to manage the economics of observability data. Explore Splunk's data tiering methodology research to learn more about how you can implement a framework that builds a solid foundation. Splunk understands data economics, knows that not all data is created equal and automatically selects the proper standards — helping you to save time, money, resources and effort.

Learn more

23-407754-Splunk-The Economics of Observability Data-EB-104

splunk>