



Transforming the Economics of Log Management

WITH NEXT-GEN CLOUD DATA PLATFORMS

WHITE PAPER

For Observability Use Cases





Observability Solutions Need a New Data Storage Strategy

The high cost of traditional observability solutions and an unpredictable macroeconomic outlook are forcing business leaders to scrutinize their annual cloud spend. But enterprises and observability platforms face a formidable challenge. How is it possible to bring observability costs down when the volume of log data has increased 5x over the last three years on average?¹

Compounding the issue, major observability platforms and software solutions are built on data management systems that aren't cost-effective for log data at scale. Unfortunately, without changing the underlying data storage layer, there is no way to avoid passing these costs onto customers.

To solve the problems of data at scale while remaining cost-effective, modern observability platforms must embrace new, cloud-native approaches. These approaches include many if not all of the following design choices: decoupled storage that uses low-cost cloud object stores, stateless ingest, query, and storage components, truly elastic scaling, and advanced data compression.

OVERVIEW

In this report, you'll learn about the following:

- Why today's observability solutions costs too much
- Why customers must throw away data to control costs
- How data-hungry use cases like AI and cybersecurity are driving companies to find new, low-cost observability options
- How next-generation observability platforms are transforming the economics of observability, making it possible to keep more data for longer

“When two great forces oppose each other, the victory will go to the one that knows how to yield.”

– Lao Tzu, Tao Te Ching

INTRODUCTION

Right now in the world of cloud infrastructure, two massive forces are on a collision course with enterprise users caught in the middle. On the one hand, the amount of log data continues to grow at vertiginous rates. This is high-value data that provides insights crucial to doing business in the cloud—the data at the core of all observability use cases.

This avalanche of data is too expensive to ingest and store with many leading observability technologies that use inefficient and outdated data storage architectures.

Tech leaders are trapped in the middle and must choose between reducing costs or throwing away data. According to a 2022 IDC poll, 63% of surveyed CIOs planned to reduce their IT spend or overall IT budgets.² And 91% of DevOps and Site Reliability Engineering (SRE) respondents in a 2023 survey said their leadership will more closely scrutinize observability and monitoring costs.³

At the same time, the volume of incoming log data is accelerating. According to the same survey of DevOps and SRE professionals, log data has grown 5x over the past three years on average, and nearly a quarter of respondents (22%) experienced a growth rate of 10x or higher.⁴

Observability is more business-critical than it’s ever been—but how are companies supposed to reduce their observability spend when they are ingesting and monitoring more data?

Something must give, and yet the volume of data will continue to grow. So the approach observability technologies take to data storage must change.

Today's Observability Costs Too Much

In 2022, reports surfaced that a leading observability provider had a customer (Coinbase) with an eye-watering \$65 million yearly bill in 2021.⁵ Challenging economic conditions meant that Coinbase had to renegotiate their contract in 2022. While Coinbase's bill was especially exorbitant, many other enterprises are dealing with the same issue—the combination of more difficult economic conditions combined with sky-rocketing observability bills.

According to Charity Majors, the CTO of Honeycomb, observability should be 20-30% of infrastructure spending.⁶ But does that estimate actually reflect the importance of observability? Or is it an attempt to justify the high costs of many observability platforms?

Costs are continuing to go up, yet many professionals aren't satisfied with the return on investment.

The survey of DevOps and Site Reliability Engineers (SRE) professionals suggests that from a qualitative perspective, professionals feel that costs are simply too high. 84% believe their observability solutions aren't currently delivering enough of a return on investment.⁷ Ironically, according to another survey, 72% of companies expected to increase or maintain their observability budgets in 2023.⁸ Costs are continuing to go up, yet many professionals aren't satisfied with the return on investment.

One area that has seen especially high costs is security information and event management (SIEM), with costs as high as \$18 million per year on average.⁹ Not surprisingly, 43% of security practitioners surveyed in a report felt they were paying too much for their SIEM solution. And less than 20% of those practitioners believed that their current SIEM solution provides value that exceeds the cost of the solution.¹⁰

Long-standing observability platforms like Splunk and Datadog have done an admirable job creating a wide range of features for monitoring applications. In fact, the observability industry would not exist without pioneers that made features like APM (application performance monitoring), RUM (real user monitoring), and instrumentation (the process of using code to collect telemetry data) possible. These platforms deserve kudos for their mature user interfaces, rich feature sets, and extensive integrations, all which are designed to provide observability in the fullest sense of the word—visibility into every part of your application.

However, observability is only effective if it's affordable, and addressing the cost of big data is much more challenging for incumbent observability platforms. Incumbent platforms are enterprises that have built mature observability offerings, but now have to manage legacy data layers. While these platforms offer

observability at scale, they are much more expensive than next-generation observability platforms built on decoupled, stateless architecture that use cloud-native tools like Kubernetes. Incumbent observability platforms typically rely on stateful architecture that uses expensive locally attached storage, which drives up both compute and storage costs. If you are using a SaaS platform, you will not have visibility into this issue, but your platform will still pass the costs on to you.

For observability solutions you run on your own infrastructure, tightly coupled components are difficult to scale up and down. As a result, you must overprovision and permanently scale up and maintain capacity in order to meet peak demand and compensate for less efficient system components. This leaves resources idle, but you still have to pay for them. Furthermore, for optimal query performance, a platform may use expensive solid-state drives (SSDs) for storage and large memory pools for compute, which can push costs even higher. And while these platforms do provide a wide range of features and integrations, the cost of developing these features is passed on to customers—even if they don't use all of the features.

To see how last-generation big data processing and storage architectures lead to high costs, consider incumbent observability platforms like Splunk, DataDog, New Relic, and Elastic. While these platforms offer all the benefits that you'd expect from mature software solutions, including extensive integrations and features, their underlying infrastructure was not originally built with technology like Kubernetes and cloud object storage in mind. They were not architected to handle current volumes of log data, let alone what's expected in the future.

Splunk was first released in 2003, New Relic in 2008, DataDog in 2010, and Elastic in 2012. Meanwhile, in 2010, the entire internet generated approximately 2 zettabytes of data.¹¹ (A zettabyte is a trillion gigabytes.) By 2022, that number was 97 zettabytes—almost a 50x increase in just over a decade. And that number is expected to double again by 2025.

It isn't just the increased data volume but also the explosion in the sheer number of systems that require observation that are driving demand for a new approach to observability. Kubernetes adoption alone has grown 127% per year, with 83% of enterprises in one survey running ten or more clusters that generate log data that must be stored and analyzed.¹²

The log data is coming from many directions at once, including microservices-based and serverless architectures, security monitoring tools, and applications for end-users. This data requires increasingly sophisticated observability solutions to ingest and combine data from many sources, analyze that data, and quickly find and troubleshoot issues. While incumbent observability solutions were built to scale over the internet of 2010, they are having challenges dealing with the demands of big data today.

Cost Control Versus Keeping Your Data: A False Choice

In order to manage costs, customers of observability platforms are often forced to limit the amount of observability data they retain. In fact, 98% of respondents in the DevOps and SRE survey attempt to remedy cost issues by limiting the logs that their observability platform ingests.¹³

In a two-part series on reducing observability costs in The New Stack, four solutions were presented for reducing costs: downsampling, limiting data retention, limiting dimensionality, and aggregating data.^{14 15}

These solutions involve collecting some, but not all, of the observability data that a system generates.

- **Downsampling** involves collecting and ingesting only a portion (often randomized) of incoming data.
- **Limiting data retention** means that data is only stored for a short period of time (typically from 7 to 30 days) before it's discarded entirely or moved to cold storage that's difficult to access.
- **Limiting dimensionality** means incoming data isn't as granular because some attributes are discarded.
- **Aggregating data** means that data is summarized in a more general way, such as through a metric, so the remaining data loses its specificity and much of its usefulness.

Log data is like evidence at a crime scene. It loses its value when sampled, aggregated, or otherwise discarded.

Log data is valuable precisely because it can provide a detailed explanation of what happened at a point in time. It is like evidence at a crime scene. It loses its value when sampled, aggregated, or otherwise discarded. All of these approaches, while presented as cost-saving measures, limit the effectiveness of your log data and, perhaps more importantly, reinforce a false dilemma where you must choose between keeping your data versus constraining costs. And they also create additional issues downstream.

For example, short retention periods in always-available "hot" storage means that data is either thrown away or quickly moved to cold storage where it needs to be rehydrated for access. According to a report from Splunk, nearly 50% of all data is dark data—data that can't be readily accessed or analyzed.¹⁶ A fifth of respondents said that more than 70% of their data is dark. Because customers can efficiently query their

data for only 7 to 30 days, it quickly becomes inaccessible. While cold storage may be cheaper, you pay the hidden cost of lost business insights and risk damaging your brand if you can't quickly resolve performance and security issues.

Artificial intelligence is often presented as a potential solution for the dark data problem. According to the Splunk whitepaper, "Dark data provides an enormous, untapped resource of information that AI can analyze. And AI-powered analytics tools can help make dark data ready for analysis on a scale that would be impossible with current methods."¹⁷

While this is true, AI models need always-available hot storage for training runs and data analysis. To fully take advantage of AI tools such as predictive analysis, you need as much training data as possible. If you're throwing data away or moving it into cold storage, it's much more challenging to maximize the benefits of AI.

The problem of lost and inaccessible data can be anxiety-inducing for teams, from the CTO that needs to reduce cloud spend to the engineering managers and teams, including DevOps and SREs, that must find and fix issues with potentially incomplete data. Teams may also have to deal with unnecessary friction related to this dilemma when the best solution is to adopt new technologies that transform the underlying economics of observability altogether.

Transforming the Economics of Observability

For enterprises building either customer-facing observability products or internal tools, the core challenge of cost-effective observability is at the heart of the platform itself: the data layer and the software used to process data.

Regardless of whether the offering is B2B or internal, platforms must be able to handle:

- Year-over-year double-digit growth in data volume.
- Long-term hot data retention windows that allow teams to analyze both current and historical data—and also satisfy the growing AI-driven demand for data.
- The need to combine disparate sources of streaming data into single, actionable views so teams can better understand their data in distributed microservices environments.

They must accomplish all this while reducing observability spend relative to overall cloud infrastructure budgets.

Incumbent observability platforms and software can make incremental changes to their current architectures to temporarily address these challenges, but to successfully transform the economics of observability—and to be future leaders in the observability space—these enterprises must move to cloud-native, decoupled architectures built specifically to process streaming log data at petabyte scale.

A new generation of observability platforms is maximizing the benefits of low-cost object storage, allowing customers to store data on their own virtual private cloud (VPCs) for greater cost savings and control, using improved compression algorithms to minimize the size of their data footprint, and using cloud-native, decoupled architecture to separate storage from compute for maximum scalability. Let's take a look at how each of these architectural shifts changes the underlying economics of observability, making long-term retention at scale affordable for enterprises.

- Object storage is considerably less expensive than SSDs or SSD-backed cloud block storage such as AWS Elastic Block Storage. Cost-effective storage allows enterprises to retain data longer and eliminates the need for tiered data storage, ensuring data is always available for teams and AI applications.
- Enterprises can benefit from additional storage and compute cost savings with their cloud providers by storing data on their own VPCs. Customers pay for their own cloud infrastructure, not the more expensive legacy data layers that mature observability platforms typically rely on. In addition to reducing costs, this also makes pricing more transparent.
- With cloud-native, decoupled architecture, storage and compute-intensive query and ingest subsystems are separate. Enterprises can scale compute up for peak events and down to zero during off-peak times. This prevents overprovisioning and also creates greater cost savings when demand is low.
- Advanced compression algorithms can reduce an enterprise's data footprint by 90% or more, significantly reducing storage costs.

With these features, next-generation observability tools are changing the economics of ingesting, analyzing, and storing log data. But it isn't just about reducing costs. Eliminating cost constraints combined with the ability to handle petabyte-scale workloads creates an opportunity to build entirely new classes of solutions for AI-enabled products, MLOps, security and compliance, ad tech, and media delivery.

By adopting technologies that change the underlying economics to make big data cost-effective, enterprises get both affordable observability and the freedom to focus their budget and mental capital on business-critical decisions like new features.

With these tools, enterprises will no longer have to worry about anxiety-inducing data retention policies and budget-busting observability bills. By adopting technologies that change the underlying economics to make big data cost-effective, enterprises get both affordable observability and the freedom to focus their budget and mental capital on business-critical decisions like new features. Finally, by bringing dark data to light and ensuring that all your data is accessible, these solutions can help answer business questions that couldn't be answered before, providing leaders and teams the information and confidence they need to make data-driven decisions that lead to future organizational success.

Hydrolix: Built to Power Next-Gen Observability

Hydrolix is designed with the future of observability in mind. It can process and retain log data at petabyte scale at a fraction of the cost of incumbent technologies. You can use Hydrolix to build B2B observability products, develop in-house observability solutions, or manage your log data at scale, all while maintaining a total cost of ownership (TCO) that is typically at least four times less expensive than other observability solutions.

Effective cost for 1 TB/ingest and 15 months of data retention:^{18 19 20 21 22}

Company	Effective Cost per GB
Hydrolix.....	\$0.32
New Relic	\$1.00
Splunk (max 30 days)	\$2.19
Datadog.....	\$34.52
ElasticSearch.....	\$63.07

Hydrolix incorporates the core principles of modern cloud-native infrastructure and stateless, decoupled architecture outlined above to provide the following benefits:

- Data ingest, storage, and analysis designed for petabyte-scale use cases.
- Hot storage performance at cold storage prices for AI, security, and observability use cases.
- High-density compression algorithms that reduce your data footprint—and your storage costs.
- The ability to keep all your data in your own VPC, allowing you to control your data while reducing costs.

In building Hydrolix, we've tried to take Lao Tzu's advice to heart. We've built a log data processing engine that won't break under the flood of data. Instead, what must yield is the way log data is stored and with it, the cost of observability. If you share our conviction that log data will continue to increase and, contrary to the law of supply and demand, become more valuable, then we would love to talk.

Visit us at hydrolix.io to learn more about how Hydrolix is powering companies at the forefront of observability. Or email us at info@hydrolix.io to arrange a demo.

Endnotes

- 1 Edge Delta. (n.d.). *Charting Observability 2023*. Retrieved December 11, 2023, from <https://observability.edgedelta.com/hubfs/Collateral/Charting-Observability-2023.pdf>
- 2 Villars, R., & Carter, P. (2022, July 22). *The First Potential Recession in the 'As-a-Service' Technology World*. IDC Blog. Retrieved December 11, 2023, from <https://blogs.idc.com/2022/07/22/the-first-potential-recession-in-the-as-a-service-technology-world/>
- 3 *ROI from Observability and Monitoring Initiatives Not Keeping Pace with Rising Costs*. (2023, September 14). PR Newswire. Retrieved December 11, 2023, from <https://www.prnewswire.com/news-releases/roi-from-observability-and-monitoring-initiatives-not-keeping-pace-with-rising-costs-301927290.html>
- 4 Edge Delta. (n.d.). *Charting Observability 2023*. Retrieved December 11, 2023, from <https://observability.edgedelta.com/hubfs/Collateral/Charting-Observability-2023.pdf>
- 5 Orosz, G. (2023, May 12). *Datadog's \$65M/year customer mystery solved*. The Pragmatic Engineer. Retrieved December 11, 2023, from <https://blog.pragmaticengineer.com/datadog-65m-year-customer-mystery/>
- 6 Majors, C. (2022, August 18). *How Much Should My Observability Stack Cost?* Honeycomb. Retrieved December 11, 2023, from <https://www.honeycomb.io/blog/how-much-should-my-observability-stack-cost>
- 7 *ROI from Observability and Monitoring Initiatives Not Keeping Pace with Rising Costs*. (2023, September 14). PR Newswire. Retrieved December 11, 2023, from <https://www.prnewswire.com/news-releases/roi-from-observability-and-monitoring-initiatives-not-keeping-pace-with-rising-costs-301927290.html>
- 8 New Relic. (2023, October 30). *2022 Observability Forecast*. Retrieved December 11, 2023, from <https://newrelic.com/observability-forecast/2022/about-this-report>
- 9 Schoenbaum, D. (2022, November 28). *The Average SIEM Deployment Costs \$18M Annually...Clearly, Its time for a change!* Dan Schoenbaum. Retrieved December 11, 2023, from <https://schoenbaum.medium.com/the-average-siem-deployment-costs-18m-annually-cf576f6c740d>
- 10 Hope, A. (2021, September 9). *43% Of Security Practitioners Believe They Are Overpaying for Their SIEM*. CPO Magazine. Retrieved December 11, 2023, from <https://www.cpomagazine.com/cyber-security/43-of-security-practitioners-believe-they-are-overpaying-for-their-siem/>
- 11 Duarte, F. (2023, December 13). *Amount of Data Created Daily (2024)*. Exploding Topics. Retrieved December 11, 2023, from <https://explodingtopics.com/blog/data-generated-per-day>
- 12 Kelly, A. (2023, January 16). *Kubernetes in the wild report 2023*. Dynatrace. Retrieved December 11, 2023, from <https://www.dynatrace.com/news/blog/kubernetes-in-the-wild-2023/>
- 13 *ROI from Observability and Monitoring Initiatives Not Keeping Pace with Rising Costs*. (2023, September 14). PR Newswire. Retrieved December 11, 2023, from <https://www.prnewswire.com/news-releases/roi-from-observability-and-monitoring-initiatives-not-keeping-pace-with-rising-costs-301927290.html>
- 14 Mitchell, A. (2023, September 5). *Top Ways to Reduce Your Observability Costs: Part 1*. The New Stack. Retrieved December 11, 2023, from <https://thenewstack.io/top-ways-to-reduce-your-observability-costs-part-1/>
- 15 Mitchell, A. (2023, September 12). *Top Ways to Reduce Your Observability Costs: Part 2*. The New Stack. Retrieved December 11, 2023, from <https://thenewstack.io/top-ways-to-reduce-your-observability-costs-part-2/>

- 16 Splunk. (2022, October 2). *The State of Dark Data*. Retrieved December 11, 2023, from https://www.splunk.com/content/dam/splunk2/en_us/gated/white-paper/the-state-of-dark-data.pdf
- 17 Splunk. (2022, October 2). *The State of Dark Data*. Retrieved December 11, 2023, from https://www.splunk.com/content/dam/splunk2/en_us/gated/white-paper/the-state-of-dark-data.pdf
- 18 Hydrolix. (n.d.). Retrieved December 11, 2023, from <https://hydrolix.io/pricing/>
- 19 New Relic. (n.d.). Retrieved December 11, 2023, from <https://newrelic.com/pricing>
- 20 Datadog. (n.d.). Retrieved December 11, 2023, from <https://www.datadoghq.com/pricing/?product=log-management#products>
- 21 AWS Marketplace: *Splunk Enterprise*. (n.d.). AWS. Retrieved December 11, 2023, from <https://aws.amazon.com/marketplace/pp/prodview-l6oos72bsyaks>
- 22 Elastic. (n.d.). *Official Elasticsearch Pricing: Elastic Cloud, Managed Elasticsearch*. Retrieved December 18, 2023, from <https://www.elastic.co/pricing/>