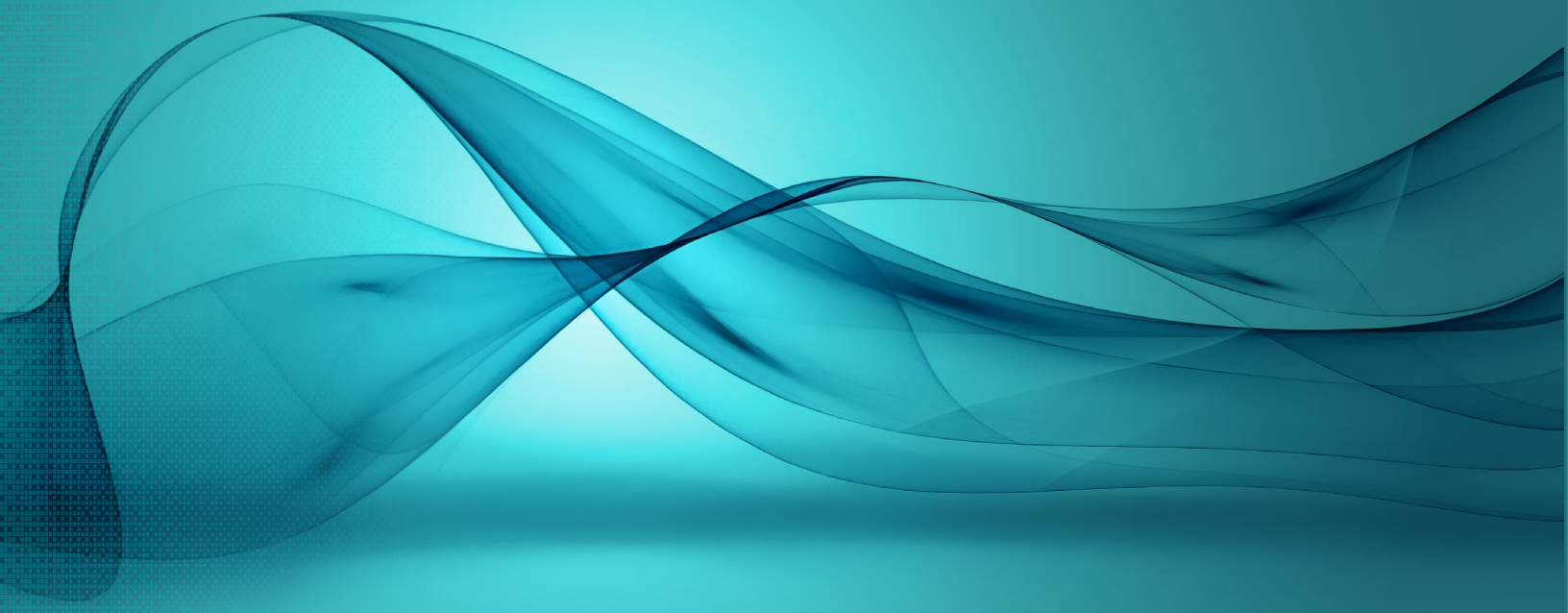
Cribl

>WHITE PAPER_

What is Observability?



>WHITE PAPER_

What is Observability?

Observability allows IT and Security teams to understand applications and infrastructure from the data they produce.

Over the last two years, the increasing complexity of modern distributed systems and application architectures has highlighted the limits of legacy monitoring approaches. Legacy monitoring remains fixated on collecting and reporting errors, restricting its effectiveness in today's dynamic and ephemeral environments. Observability takes a new approach, allowing teams to interrogate system behavior without the limits imposed by legacy methods and products. Coping with the data volumes and variety of observability data, as well as the costs, requires a new approach to managing data.

THE CHALLENGE

Monitoring helps teams answer known questions about their infrastructure and environment, but it doesn't give insight into new signals or opportunities.

THE SOLUTION

On the other hand, observability allows staff to understand new or unexpected behaviors in applications and infrastructure from the data they produce.

THE BENEFITS

- Get insights into application and infrastructure behaviors you didn't know existed.
- Discover unknown events and develop a plan of action.
- Complement your existing monitoring practices for full insight into your environment.

Defining observability.

There are as many definitions of observability as there are vendors in the market today. The definition we like at Cribl comes from Gartner:

> **"Observability is the characteristic of software and systems that allows them to be "seen" and allows questions about their behavior to be answered."**
>
> — From Innovation Insight for Observability

That seems straightforward enough, but this simple definition has created confusion among site reliability engineers (SREs), SecOps, and ITOps teams. Many conflate observability with its older cousin, monitoring. This misses the point. Observability isn't about alerts and dashboards. Those are the things you already know. Observability is about discovering things you didn't know.

Another way to think about the differences between observability and monitoring is with a data infrastructure analogy. A monitoring platform is like a data warehouse. Well-known, well-understood data is ingested into your platform of choice and shaped to answer the questions you know you want to ask. That's what a monitoring platform is good at: answering the known questions.

Observability is more like an exploration environment, like a data lake. Data lakes collect data from across the organization and skilled users explore stored data for new signals or opportunities. When a promising opportunity is found, the data is refined and optimized, then used in a more consumable platform, like a data warehouse.

The need for observability and monitoring.

The upswell of interest in observability continues driving confusion in the monitoring space. IT teams question if they need observability and how existing monitoring tools will work in this new concept. Some of the questions we hear from our customers are:

- Does observability replace monitoring?
- How do observability and monitoring work together?
- What new tools do I need to take advantage of observability?

The reality is you need both observability and monitoring for full insight into your environment, including applications and infrastructure. The easiest way to think about these two concepts is the popular OODA Loop. The OODA Loop has been used for years in domains as diverse as military strategy to IT operations. If you're unfamiliar with OODA, it stands for Observe, Orient, Decide, and Act. It's a good way to think about how observability and monitoring are complementary to each other.

You need both observability and monitoring for full insight into your environment, including applications and infrastructure.

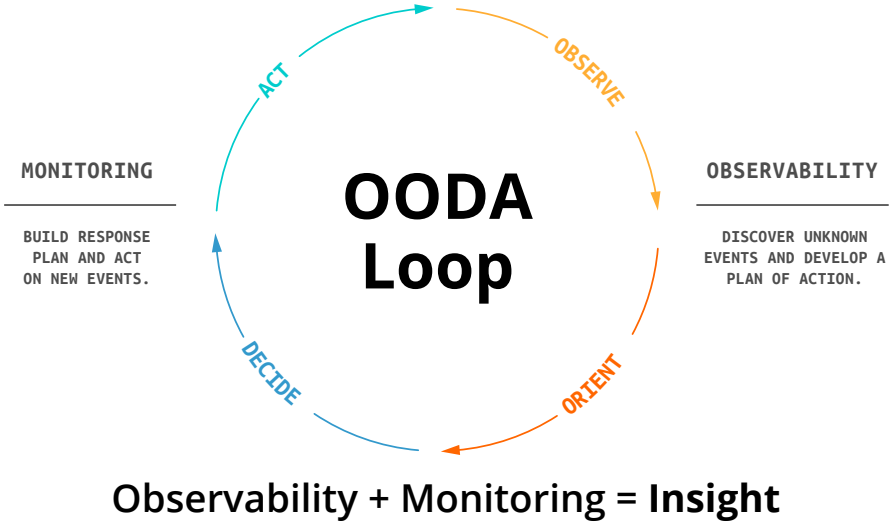


Figure 01: The OODA Loop.

On the right side of this Figure 1, you have Observe and Orient. This is where the observability side of the equation comes in. The left side, Decide and Act, is the monitoring part of the equation. Think about the two halves of the equation like this: Observability lets you discover new or unexpected signals in your environment. These are the things you're not monitoring for yet — you don't even know they exist! This is what observability gives you: The ability to discover and understand behavior of applications and infrastructure from the data these things emit. Observability is about discovery.

Monitoring, the second half of the equation, is about doing. This is where Decide and Act come in. Once you know the data indicates a performance challenge or a potential security breach, you can configure your monitoring systems for those signals and automate the types of responses that are most suitable for these new conditions.

While the concepts of observability and monitoring are closely related, there are differences. The biggest difference starts with data.

The diversity of observability data.

Traditional monitoring platforms take an exclusive approach to the data they use, giving them a limited view of the environment they monitor. These platforms primarily use data from dedicated software agents deployed across applications and infrastructure. Agents only send data to one place: the target monitoring platform. Most enterprises have twenty to thirty monitoring tools, each with their own dedicated agents, resulting in disconnected silos of information. This limits the effectiveness of monitoring solutions.

Observability takes the opposite approach. With an inclusive approach to all sources of data, observability tools can deliver better value for discovery. Today, the focus is on four distinct data types: events, logs, metrics, and traces, commonly shortened to MELT.

In the near future, Cribl believes observability platforms will consume a much more diverse array of data sources, including configuration management data, dependency maps, and data from data marts, lakes, and warehouses. Similarly, these diverse sources will be delivered to a range of new and evolving destinations.

Connecting these sources to destinations in an intelligent way is the first observability challenge. This is where the observability pipeline comes in.

With an inclusive approach to all sources of data, observability tools can deliver better value for discovery.

Abstracting sources from destinations with an observability pipeline.

An observability pipeline is a strategic control layer positioned between the various sources of data, like networks, servers, applications, and software agents, and the multiple destinations in today's IT and SecOps environments. Instead of relying on siloed point-to-point connections, an observability pipeline centralizes all of your observability data processing, giving your teams full control over every aspect of your data.

From a data management perspective, observability pipelines allow you to:

- Filter out redundant data to improve performance and lower the cost of ingest-based destination platforms.
- Enrich data with additional context, like GeoIP, for improved downstream analysis and compliance with various data privacy laws.
- Redact sensitive data to comply with data governance requirements.
- Route data from one source to multiple destinations, eliminating the need to deploy new agents alongside deploying a new service.
- Send full-fidelity data to low-cost storage and replay it when needed.

Abstracting the sources and destinations of observability data offers massive benefits to IT and SecOps teams, including:

- Providing a single point for governing data and applying consistent rules for data redaction, access control, and sharing.
- Reducing the amount of redundant data flowing into downstream systems like logging analytics, SIEM, and SOAR platforms.
- Accelerating onboarding new tools by sharing data from one source with multiple destinations.

As the need for more data becomes apparent, another challenge must be overcome, and that's the need for dynamic data tiering based on the changing value of data.

The need for dynamic, value-based data tiering.

On average, IT and Security teams are collecting 28% more data year over year. While much of that data lands in high cost destinations, the reality is most data is never explored or analyzed. It's kept just in case - just in case an application goes down, or just in case a breach occurs. Until you need that data, its value is essentially zero but you're storing it in a high cost data tier. That drives up observability costs. We need a new approach to how we retain data.

By adopting a data philosophy of aligning where data is stored aligned its value and usage, performance optimized data can be delivered to the range of analytics, monitoring, and cybersecurity tools in use, providing the most advantageous data product for those platforms. Full fidelity data, meanwhile, is readily delivered to cost optimized object storage. This full fidelity storage is often required for exploratory or compliance use cases. This tiering strategy allows users to put data in the right location for its desired outcomes and use, aligned to the value of the data.

With data stored in a tier relative to its value, we next have to think about governing, sharing and accessing data across teams and across companies.

Instead of relying on siloed point-to-point connections, an observability pipeline centralizes all of your observability data processing, giving your teams full control over every aspect of your data.

Distributing data access and governance.

As data is stored across value-based tiers, we've shifted from forcing users to centralize data in the most expensive location. Next, we need a way to access data wherever it resides, whether that is at an endpoint, in a relational database, within APIs, or low-cost object storage. What we need is a way to unify data at the query tier instead of the storage tier, giving users a consistent, immutable view of their data assets by adopting schema-on-need instead of the schema-on-write or schema-on-read approaches taken by legacy platforms.

Schema-on-need is a new approach to providing insights for IT and security data combining the benefits of schema-on-read and schema-on-write. This eliminates the need to predefine a rigid schema, while still providing the performance, quality, and governance benefits of working with a schema. Unlike schema-on-write, where all data is structured and typed before it is stored, schema-on-need defers working with a schema until needed. This critical feature of the data engine examines how data is accessed and used, then applies the optimal schema based on those access patterns. This approach minimizes the effort and complexity associated with defining schemas for all data upfront.

Additionally, the query language meets users where they are through an open and familiar pipe-delimited format applicable to any queried source. This removes the need for mastery of multiple languages. And as one part of the data engine, results can be readily sent to any destination through the data engine's observability pipeline capability.

An observability data management strategy for modern IT and Security Teams.

Whether you're on the IT side or the Security side, observability — and the data required to achieve it — is a challenge. It is clear that a new approach is needed to manage and use observability data. What's needed is a new strategy for managing data aligning to the business, regulatory, and financial realities enterprises find themselves grappling with not only today, but also into the future. We need new and flexible observability data management strategies.

Data management strategies cannot remain fixed. They must be adaptable to unpredictable futures. IT and security teams need a unified, composable, and flexible data management infrastructure as a basis of their strategy for collecting, processing, and accessing data at scale.

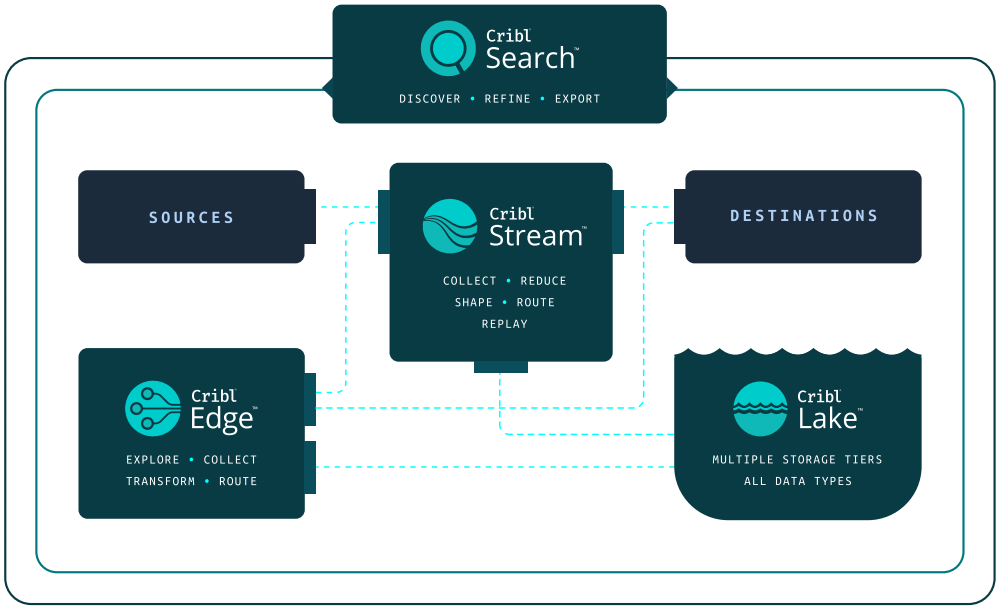


Figure 02: Cribl suite.

Observability gives you the opportunity to discover and understand your dynamic environments in near real time.

The elements of an observability data strategy removes the compromises present in both incumbent and insurgent solutions by offering essential capabilities composable into architectures optimized for a given environment and use cases. The core capabilities of a modern data infrastructure consist of:

- Endpoint monitoring and management with rich data collection and processing capabilities, complete with data access, control, and ability to search distributed data sets.
- Robust pipeline features to route, enrich, govern, and control data at scale.
- Distributed data sharing and access with strong governance.
- Supported by an opinionated but flexible and automated view on data storage, tiering, and metadata management, built on open data formats.

When composed, these capabilities create a data engine able to power the needs of IT and security teams. By adopting a composable observability data infrastructure designed for the unique and diverse needs of today's IT and security teams, organizations can navigate data challenges efficiently. The answer lies in the adoption of a data engine tailored to modern IT and security operations that provides a unified, composable data management platform for today and into the future.

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

WP-0001-EN-2-0624