Cribl

>WHITE PAPER_

Modernizing data management for IT and security.

Cribl

>WHITE PAPER_

Modernizing data management for IT and security.

Evolving demands placed on IT and Security teams are driving a new architecture for how observability data is captured, curated, and queried. This new architecture provides flexibility and control while managing the costs of increasing data volumes.

Key takeaways

- Increasing regulatory scrutiny, expanding data sharing and access requirements, and exploding data volumes are driving a need for a modern approach to data management for IT and security teams.
- Coping with evolving demands forces IT and security teams into either legacy fit-for-purpose tools, with their high costs and vendor lock-in, or labor intensive general purpose data platforms.
- Current legacy tooling forces IT and security teams into a one-size-fits-all approach to data, requiring centralized data storage that limits flexibility and drives up costs.
- Telemetry data — the universe of metrics, events, logs, and traces that makes IT and security possible — has unique characteristics from transactional data, notably in its staggering volume, unstructured nature, and broad variety of formats.

Current legacy tooling forces IT and security teams into a one-size-fits-all approach to data, requiring centralized data storage that limits flexibility and drives up costs.

Recommendations

- Scope your data management challenges by forecasting your enterprise's cybersecurity and observability data collection and analysis needs over a 12, 24, and 36 month horizon and contrast that with your forecasted IT budget growth over the same time period.
- Search for opportunities to consolidate your cybersecurity and observability tools and platforms using the capabilities around a unified data approach leveraging open storage and access formats: from endpoint, to pipeline, to tiered storage, and search.
- Build a data management strategy that efficiently delivers data into tiered storage based on timeliness requirements for operational versus exploratory use cases, while also accounting for retention requirements for regulatory and compliance needs.

The disconnect between operational vs. organizational needs

Today's IT and security teams are coping with competing internal and external forces. From the outside, these forces range from an increasing number of attacks, to an explosion in the number and complexity of endpoints and deployment environments, and a need to ensure application performance and uptime. Internally, these teams are tasked with retaining ever larger amounts of data for compliance and regulatory purposes, sharing data across diverse teams and geographies, and dealing with legacy platforms designed for previous versions of IT and security operations that simply no longer exist.

The tailwind driving each of these forces is the same - the ever-increasing volumes of generated data. Recent studies have found that enterprise data is growing at a 28% compounded annual rate[1]. Without proportionate budget increases and storage capacity, this growth has commonly been managed by reducing the sources of data being collected or the frequency of collection, often with unintended consequences. Increasing data retention requirements mean this option is quickly being regulated out of practice and teams also have to hold onto more data, longer.

While data warehouses are essential for business intelligence and analytics, they do not have the necessary characteristics for operational use cases.

Within this data storm, teams seek shelter in various ways, each having a core disconnect with the realities of their operational and organizational needs. From infrastructure, to pricing, to staffing, the needs of the modern, data-driven organization have quickly outpaced the capabilities of traditional technology to meet them.

Disconnect #1 – New tools built on legacy technology.

Combating the rising costs of legacy solutions is pushing some teams to explore a new generation of tools built on traditional data management infrastructure, notably data warehouses. While adept at analyzing swaths of transactional data, data warehouses are a poor fit for the range of metrics, events, logs, and traces (MELT) that IT and security operations teams cope with.

- Data warehouses require a predefined structure over the data they store and process, while telemetry data is dynamic and varied. Fitting operational telemetry data into a data warehouse requires cumbersome, ongoing, data integration work. Integration delays mean data isn't available immediately - a key requirement in any operational scenario.
- For efficiency, data warehouses require all data centralized in a single location, described in a single way. Operational telemetry data resides throughout an enterprise's infrastructure, from endpoint, to data center, to cloud. Centralizing it in a single location is practically and financially impossible.
- The skills of IT and security operations teams differ from those required in data warehousing, specifically around interacting with data. SQL is the lingua franca of the data warehouse, which is a language foreign to many on the operations side of the business. Instead, operations teams lean heavily on exploratory search-based interfaces for data investigation and interrogation.

While data warehouses are essential for business intelligence and analytics, they do not have the necessary characteristics for operational use cases. The current crop of observability and security products built on top of these technologies not only fail to understand their users, they also fail to understand the realities of the data they're working with.

Disconnect #2 – Punitive incumbent pricing models.

The second disconnect is in the pricing of existing tools and platforms. Existing pricing models are outdated, typically following a few major themes, like average daily ingest, workload-based pricing, or a la carte. As data volumes increase, pricing on average daily ingest works well for vendors, but quickly puts end users on the defensive as they cut down on inbound data sources to stay within budget. When a quarter to a half of all observability data is collected in duplicate, ingest-based pricing can rapidly become unsustainable for enterprises[2].

Workload-based pricing ignores ingest-based limitations. Instead, IT and security teams incur a cost every time they access their data. This averts the limits of ingest-based pricing, but introduces another - demand. Demand, however, is harder to predict than ingestion. For example, an application outage can spike demand for accessing data as teams troubleshoot, or a security incident can result in teams digging through months of data. Both scenarios lead to skyrocketing costs under workload-based pricing.

The final pricing model - a la carte- is also driving the disconnect between vendors and their IT and security customers. In this model, simple services, like log management, are priced at pennies per gigabyte ingested. This solution is less expensive on the surface, However, simple services rarely meet the needs of diverse teams. As surrounding services are added, each with its own cost. Once all of the required services are added, it's not uncommon for enterprises to pay dollars per gigabyte under management.

As data volumes increase, pricing on average daily ingest works well for vendors, but quickly puts end users on the defensive as they cut down on inbound data sources to stay within budget.

Each of these pricing strategies greatly benefits the vendors as data volumes rise, customers' need to share and access data increases, and tiering data to required service levels becomes the norm.

Disconnect #3 – The unsustainable skills challenge.

Another disconnect highlighted above also applies to the current state of observability and cybersecurity environments - the range of products currently in use across teams. For example, security teams average nearly eighty tools in their portfolio[3]. Each product introduces its own language, nuance, and administrative overhead.

- Hiring for the right mix of skills becomes difficult. While you can train staff on the necessary skills, this delays their effectiveness and drives up costs.
- The mix of standards unevenly supported across tools introduces another challenge for staff. Data and protocol standards are a mix of legacy syslog, OpenTelemetry, Open Cybersecurity Schema Framework, and various common information models that are anything but common. Fighting with formatting data for incomplete standards wastes time and contributes to staff burnout.
- Also unworkable is the 'box of LEGOs' approach many vendors take in their products. While some audiences, such as data engineers or integrators, may thrive developing custom solutions, these are cumbersome propositions for IT and security professionals, and put enterprises one engineer away from downtime on mission critical systems.
- Automation, in the form of AIOps, is a long-promised panacea to skills challenges but has yet to realize significant gains for teams and hiring managers.

Hiring is always difficult. It is harder still when seeking a diverse range of skills. Additionally, the 'box of LEGOs' approach by vendors, although suitable for some, is impractical for sustaining a modern IT and security practice.

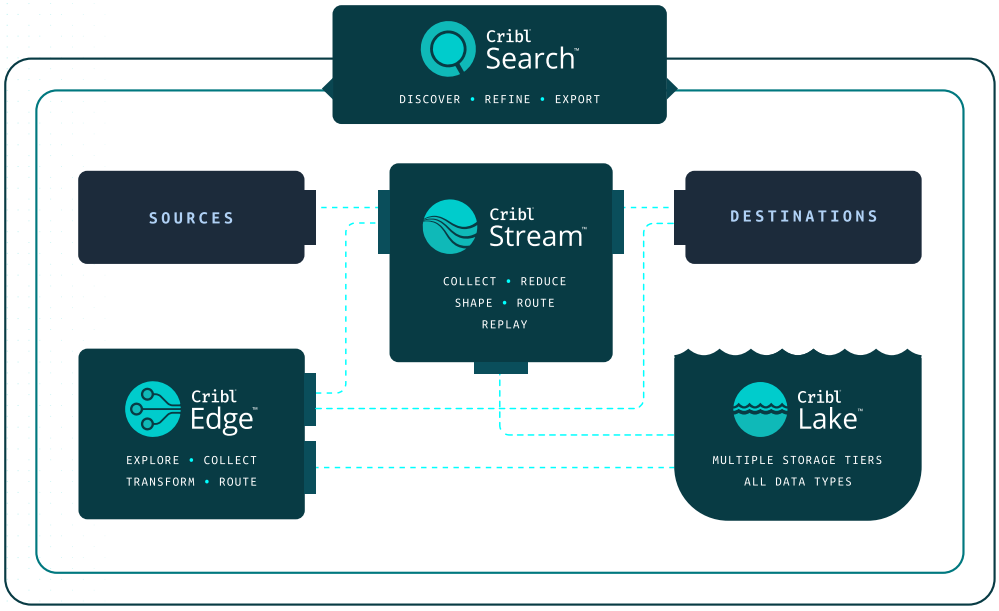
A data management strategy for modern IT and security teams.

In light of the internal and external forces working against IT and security operations teams, and the unsuccessful half-measures detailed above, it is clear that a new approach is needed to manage and use operational 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.

Addressing future requirements is difficult. After all, who can predict the future? Enterprises may enter promising new markets, or exit underperforming ones. They may acquire or be acquired, or introduce new products. Competitors previously unimagined may appear out of nowhere. Or a global pandemic may radically upend our work and personal lives, leading to shifts in every aspect of how businesses operate.

Data management strategies cannot remain fixed. They must be adaptable to these unpredictable futures.

Data management strategies cannot remain fixed. They must be adaptable to these 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.



This 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. Let's explore each of those capabilities.

Consolidated endpoint monitoring and management.

A data engine begins at the endpoint. Using an agent-based footprint, a data engine's endpoint monitoring and management allows users to collect telemetry from the application and operating system, determine what portions are valuable or require additional processing, and then choose whether to forward to downstream destinations. This endpoint capability also supports agent consolidation through rich centralized fleet management and auto configuration. As one element of a composable data engine, endpoint monitoring and management also integrates with other components, such as distributed data access and governance.

Observability pipeline.

Acting as an abstraction between the sources and destinations of operational observability data, observability pipelines allow users to route data from any source to any destination, manage data volumes at the destination, redact and filter sensitive data, in-flight enrichment, and replaying data from low-cost storage. In the context of a data engine, the observability pipeline provides the fuel and throttle for downstream components. (See "The business case for observability pipelines" for more insights.)

A data engine reflects the diverse realities of data's complex lifecycle.

Distributed data access and governance.

A data engine reflects the diverse realities of data's complex lifecycle. Rather than forcing users to centralize data before it can be accessed, the data engine allows users to access data where it resides, whether that is at the endpoint, in a relational database, APIs, or low-cost object storage. Data is unified 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.

Automated and flexible data tiering.

The data engine allows users to rationalize their data storage philosophy, aligning where data is stored with its value and usage through data tiering. 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.

Coupled with the distributed access and governance inherent in the data engine, each tier is accessible regardless of which tier data is stored within.

Your data engine action plan.

While a data engine is essentially a decomposed data infrastructure applicable to different operations teams, its ability to centralize data access and automate data tiering also streamlines the plan for adoption. Table 1 outlines the 30/60/90 day plans for IT and security leaders.

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.

	IT and security leaders.
Within 30 days.	• Forecast the increasing amount of telemetry data flowing into your enterprise, accounting for new application architectures, deployment environments, as well as legacy environments. • Inventory the range of tools and products your team uses, and map their costs to your data forecast. • Identify data engine platforms suitable for your environment, skills, and data growth needs.
Within 60 days.	• Conduct a POV of your short list of data engine platforms, evaluating them on price/performance and data accessibility. • Craft a data management strategy for telemetry data, highlighting governance, metadata management, and data sharing.
Within 90 days.	• Search for opportunities to optimize legacy tool licensing, or outright tool consolidation. • Work with your data engine vendor on use cases and future roadmap to align with ongoing needs.

Conclusion.

The escalating data challenges faced by IT and security teams won't be resolved with legacy approaches, or with new tools built on outmoded data infrastructure. By adopting a composable data infrastructure designed for the unique and diverse needs of today's IT and security teams, organizations can navigate the data storm 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 the future.

Evidence.

1. IDC. Worldwide IDC Global DataSphere Forecast, 2023-2027: It's a Distributed, Diverse, and Dynamic (3D) DataSphere. April 2023, IDC #US50554523.

2. IDC. Getting Data Collection Volume Right Is the Top Observability Challenge. November 2023, IDC #US51363123.

3. Panaseer 2022 Security Leaders Peer Report.

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-0017-EN-1-0324