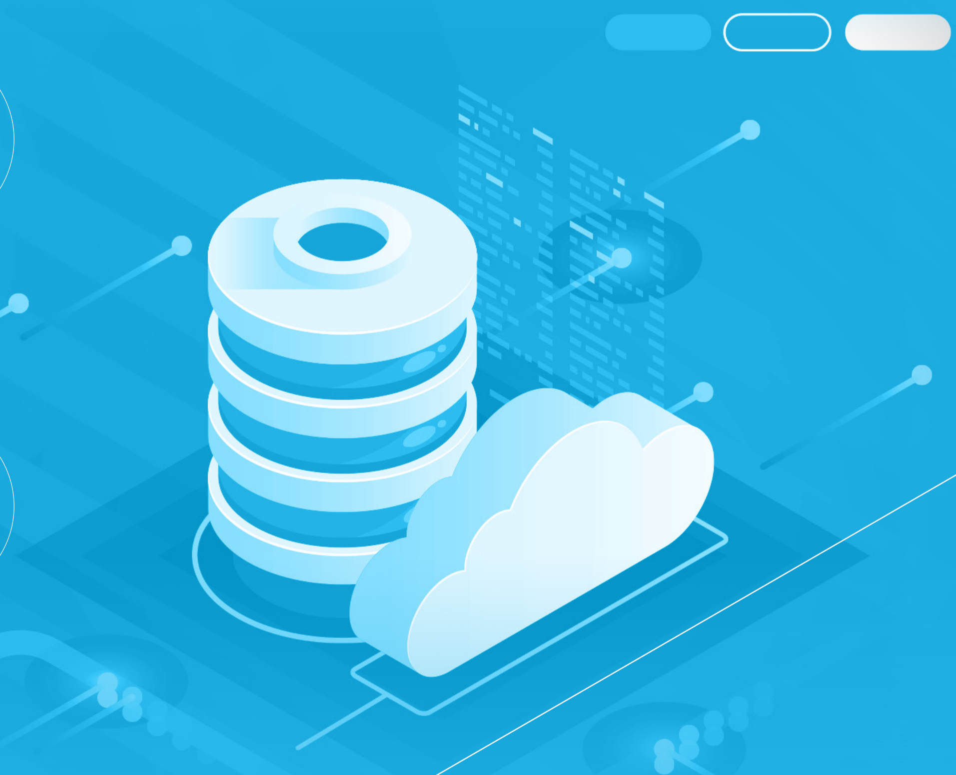
OBSERVE
BY SNOWFLAKE

THE TELEMETRY LAKEHOUSE ADVANTAGE:

HOW DATA LAKEHOUSE ARCHITECTURE IS RESHAPING OBSERVABILITY



WHITE PAPER

EXECUTIVE SUMMARY

Telemetry growth in the enterprise is breaking the economics of legacy observability. Most traditional architectures rely heavily on indexes and drive high ingestion and storage costs, tightly couple storage and compute, and force organizations into difficult tradeoffs: sampling data to control costs, dropping high-cardinality fields to manage index size, or limiting retention to days instead of months.

Organizations are rapidly adopting a new architectural model to eliminate these tradeoffs. A data lakehouse stores telemetry in commodity cloud object storage with 10x compression, separates storage from compute, and keeps all data immediately queryable without requiring tiering or rehydration. Those adopting a data lakehouse for their telemetry report up to 60% lower costs than legacy platforms, while retaining more data, for longer, without sacrificing fidelity.

Beyond economics, a data lakehouse built on open formats, such as Apache Iceberg, transforms telemetry into a portable, first-class analytical asset. Telemetry can be accessed by multiple analytical engines, enabling seamless data sharing, reducing data movement, and eliminating vendor lock-in.

SECTION 1

THE HIDDEN TAX OF LEGACY OBSERVABILITY

Legacy observability platforms were built for relatively static infrastructure: long-lived hosts, stable service topologies, and predictable data volumes. Those architectural assumptions no longer hold.

Microservices replace a handful of monoliths with thousands of deployable services, while Kubernetes and auto-scaling introduce constant churn from ephemeral workloads and dynamic infrastructure. Combined with the rise of AI agents, modern systems produce telemetry that is not just larger in scale, but significantly more complex than legacy architectures were designed to process efficiently.

The Indexing Tax

Traditional architectures achieve the performance required by observability use cases by maintaining index structures on high-performance storage, often local SSDs. That speed comes with a tax: compute and I/O amplification to build indexes on the ingested data, plus storage amplification from index structures.

The overhead from these additional structures often pushes the effective stored footprint beyond the raw data size. Replication for availability and durability then multiplies that footprint again, and because it must reside on high-performance storage, costs compound quickly. For enterprises that generate terabytes to petabytes of telemetry per day, the additional storage and compute costs become untenable.

Inefficient Data Tiering

To control costs, legacy platforms rely on data tiering. Instead of storing all telemetry on fast, high-IOPs storage, older data is moved to higher-latency tiers.

This introduces operational tradeoffs. Queries that span tiers can experience unpredictable latency, and retrieving archived data often requires rehydration, a process that can take hours and consume additional compute and storage resources.

Tiering reduces the cost of premium storage but does not eliminate storage amplification. Retention is often limited to days or weeks, constraining year-over-year analysis and historical pattern matching. As a result, engineers may find the data they need is slow to access or no longer available.

Blind Spots

Mechanisms like sampling, filtering, aggregation, and shortened data retention are often used alongside traditional observability systems to control cost and manage scale, but they can unintentionally introduce blind spots. Sampling and aggressive filtering may drop telemetry needed for troubleshooting. Heavy aggregation reduces granularity, while shorter retention windows can limit the ability to investigate trends or correlate past incidents.

Tightly Coupled Compute and Storage

Legacy systems tightly couple compute and storage, so scaling one requires scaling the other. To increase capacity, additional nodes with both compute and storage resources must be added regardless of which resource is actually constrained, resulting in inefficient utilization. This tight coupling also limits how quickly systems can scale, often forcing organizations to provision infrastructure for peak demand. As a result, resources are underutilized for much of the time, driving costs higher than necessary.

Vendor Lock-In and Data Silos

While the industry is moving toward open data collection and standardized formats, observability tools have traditionally relied on proprietary components. These proprietary data formats and collection agents increase the risk of vendor lock-in, as organizations become dependent on tooling that is costly and complex to replace.

Legacy systems built on proprietary technology also limit data sharing and often require additional data movement and replication, increasing the likelihood of isolated data silos. As a result, organizations struggle to achieve a unified observability view and to effectively correlate telemetry data across multiple tools.

Legacy Vs Data Lakehouse: The Future of Observability Architecture

	Legacy Architecture	Data Lakehouse Architecture
Ingest & Storage Efficiency	**Index-based** *Compute and storage amplification due to index structures.*	**Lakehouse-based** *10x compression on commodity cloud storage, no indexing overhead.*
Resource Utilization	**Tightly coupled compute and storage** *Inefficient resource utilization drives higher cost.*	**Compute-storage separation** *Scale elastically and independently for efficient utilization.*
Data Tiering	**Older data moved to cold tiers** *Unpredictable latencies and requires rehydration.*	**Data kept always hot** *All data remains query-able and is instantly accessible.*
Data Interoperability	LOGS METRICS TRACES **Limited data sharing** *Siloed data in proprietary formats.*	**Seamless data sharing** *Open data formats facilitate data reuse and prevent lock-in.*

SECTION 2

THE DATA LAKE ALTERNATIVE

The data lakehouse is emerging as a modern alternative to legacy observability architectures and the operational complexity they create. Coupled with the open data formats that are common in lakehouse architectures, this model addresses many of the limitations that have historically constrained observability platforms.

Eliminating the Indexing Tax

Data can be ingested into a telemetry lakehouse in its raw form, without requiring upfront indexing. Rather than relying on expensive, high-performance storage typical of legacy architectures, data is centralized in low-cost cloud object storage. Unlike traditional systems that increase storage footprint through indexing, telemetry data in a lakehouse typically achieves up to 10x compression on commodity cloud storage. As a result, the lakehouse model delivers the best economics for storing large volumes of telemetry, enabling organizations to significantly reduce storage costs while maintaining the ability to scale.

Always-Hot Data

With all telemetry in the data lakehouse, there is no need for data tiering or archival strategies that move older data into cold storage. All telemetry remains hot, fully queryable, and instantly accessible, eliminating rehydration cycles and the operational friction they create. This makes year-over-year comparisons and incident analysis over longer time windows routine rather than costly or time-consuming.

Full-Fidelity Data Retention

While legacy platforms typically retain data for only days or weeks, data lakehouse-based solutions can store months of telemetry, with some default retention periods extending to 13 months. This enables organizations to retain as much telemetry as needed without resorting to sampling, filtering, or aggregation that reduces data fidelity.

By landing raw data in the lakehouse, organizations can also defer transformation and structuring until read time. They do not need to index data or adhere to a predefined schema upfront, preserving the flexibility to materialize derived datasets from the raw data as needed for analysis.

Separation of Compute and Storage

Data lakehouse architectures decouple compute and storage, a concept popularized by Snowflake, enabling each to scale elastically and independently. Compute resources can scale up or down based on ingest or query demand, while cloud object storage provides cost-effective capacity for virtually unlimited telemetry data, ensuring efficient utilization of resources and optimizing costs.

Open Data Formats

Data lakehouse architectures are fundamentally built on open data formats, which support key principles such as reduced vendor lock-in, data reuse, and multi-engine interoperability. Modern lakehouse platforms commonly enable both reading from and writing to Apache Iceberg tables, reinforcing this open approach. Similarly, many newer observability solutions have standardized on OpenTelemetry for telemetry collection. By embracing open standards, organizations can facilitate seamless data sharing, reduce unnecessary data movement, and eliminate enterprise data silos.

SECTION 3

THE ECONOMIC CASE

The economic case for data lakehouse architecture is demonstrated most effectively through outcomes when moving from a legacy observability solution to one based on a lakehouse foundation.

What Organizations Experience

Capital One

Longer retention, lower costs at scale.

Capital One ingests over 300 TB of telemetry per day, a volume that would be cost-prohibitive with legacy indexing. With 13 months of data retention, teams can conduct historical analysis previously unattainable at this scale, while reducing observability costs by 66%.

300 TB of telemetry/day

13 months
of data retention

66% Cost Reduction

Topgolf

Consolidation and simplification

Topgolf replaced a fragmented stack of multiple tools for logs, metrics, and traces with a single platform using a common data model. Correlation that previously required manual timestamp matching is now automatic. Engineer onboarding is faster, alert management is simplified, and vendor management is streamlined.

Fragmented stack of multiple tools for logs, metrics, and traces	**Single platform** using a common data model
Manual timestamp	**Automatic timestamp**

Dialpad

Faster resolution through correlation

Dialpad reduced MTTR by consolidating log, trace, and metric correlation within a single platform. Engineers can investigate from symptoms to root causes in one interface, with service and infrastructure relationships identified automatically.

Single platform: Log, trace, and metric correlation

MTTR ↘

RelationalAI

Cost efficiency without compromise.

RelationalAI achieved improved observability at one-quarter the cost of their previous platform, while gaining vendor neutrality and enhanced data security. These savings resulted from eliminating architectural overhead such as indexing taxes and tightly coupled storage and compute, rather than reducing capability.

OBSERVABILITY ↑ | **COST** ↓

The Three Drivers of Observability Cost

- **Tool spend** is the most visible cost. Organizations report up to 60% reduction when moving to a data lakehouse architecture.
- **People costs** include engineering hours spent managing infrastructure, maintaining dashboards, and manually correlating data across siloed tools. Reducing the operational burden allows engineering effort to be redirected to product development.
- **Capability cost** refers to what organizations forgo because their architecture cannot support it, such as automated cross-signal correlation, year-over-year analysis, and the ability to analyze high-cardinality data. These opportunity costs result in longer MTTR and lower engineering productivity. Although difficult to quantify, they are often the most impactful.

SECTION 4

THE CONVERGENCE OPPORTUNITY

A data lakehouse architecture delivers value beyond cost savings and operational efficiency by enabling telemetry data to be stored and analyzed alongside enterprise data within the same lakehouse environment. By bringing operational and business data together, organizations can enrich observability workflows with relevant context.

Hard savings

Organizations eliminate data replication and movement costs and the overhead of managing separate infrastructure stacks.

New capabilities

Engineers can enhance incident response by applying business context to operational issues, while data teams can apply operational context to business decisions as needed.

Questions that become answerable include:

Which premium customers experienced latency above SLA during yesterday's outage? What was the revenue impact of last week's deployment rollback? How does infrastructure spend per service correlate with the revenue that service generates? These are not theoretical; they illustrate how observability data transitions from an operational cost center to a strategic analytical asset.

CONCLUSION

Organizations that gain a competitive advantage from observability treat telemetry as a first-class analytical asset, storing it in telemetry lakehouses rather than in indexing systems designed for a previous era.

Data lakehouses achieve this through a more scalable, modern architecture: commodity storage with 10x compression, compute-storage separation for optimal efficiency, open formats for data sharing and reuse, and the removal of the indexing tax that has defined legacy observability economics. Organizations today operate at a scale that legacy architectures cannot support. Companies like RelationalAI achieve better observability at a fraction of the cost. Teams at Topgolf and Dialpad resolve incidents more quickly because their data is unified and always accessible.

As telemetry and business data converge in the same architecture, observability becomes not just more efficient, but a critical advantage.

OBSERVE
BY SNOWFLAKE