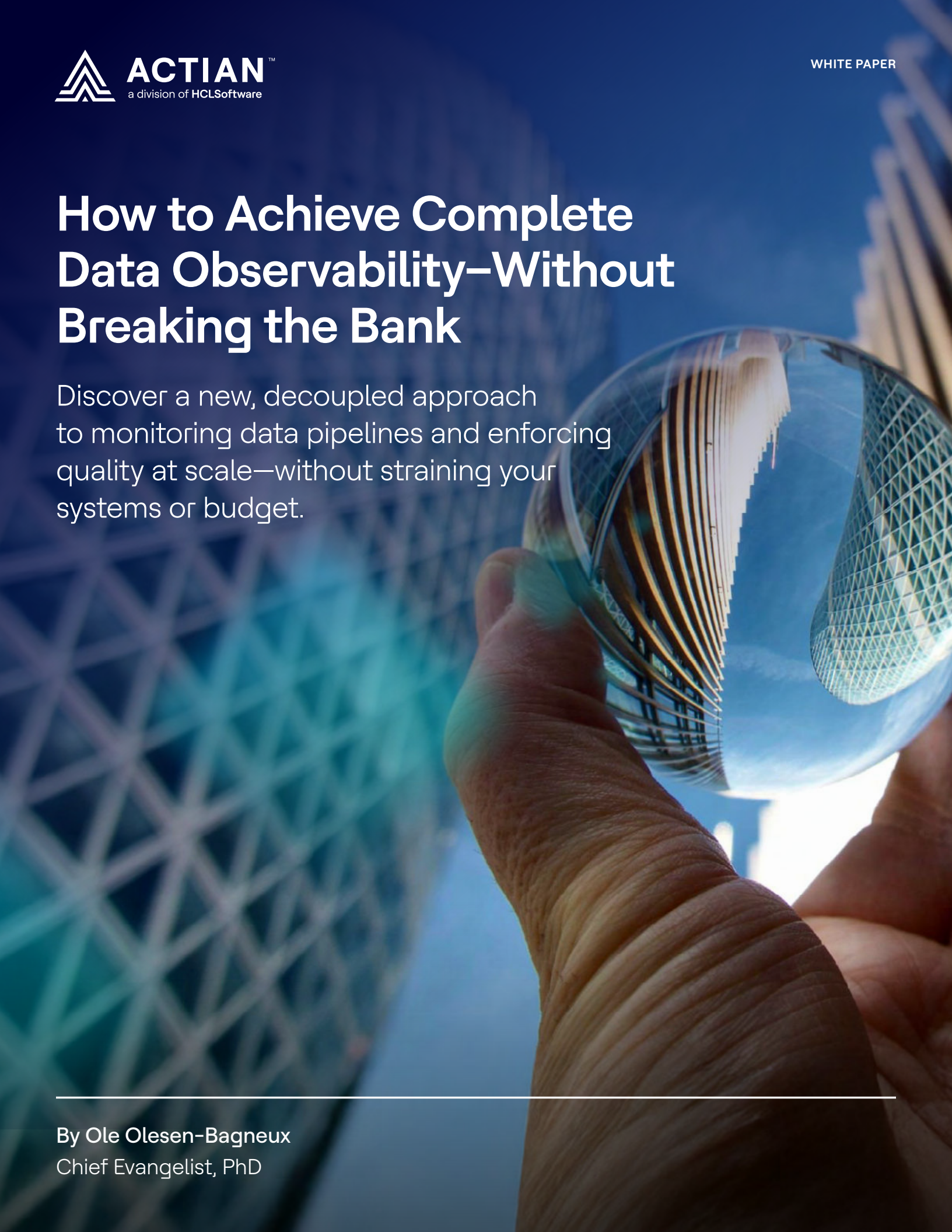
ACTIAN™
a division of HCLSoftware

WHITE PAPER

How to Achieve Complete Data Observability–Without Breaking the Bank

Discover a new, decoupled approach to monitoring data pipelines and enforcing quality at scale—without straining your systems or budget.



By Ole Olesen-Bagneux
Chief Evangelist, PhD

Table of Contents

The scale and speed of data movement today is overwhelming traditional quality controls. Actian introduces a modern, intelligent approach to data observability—bringing confidence, trust, and cost-efficiency to enterprise data operations.

About The Author

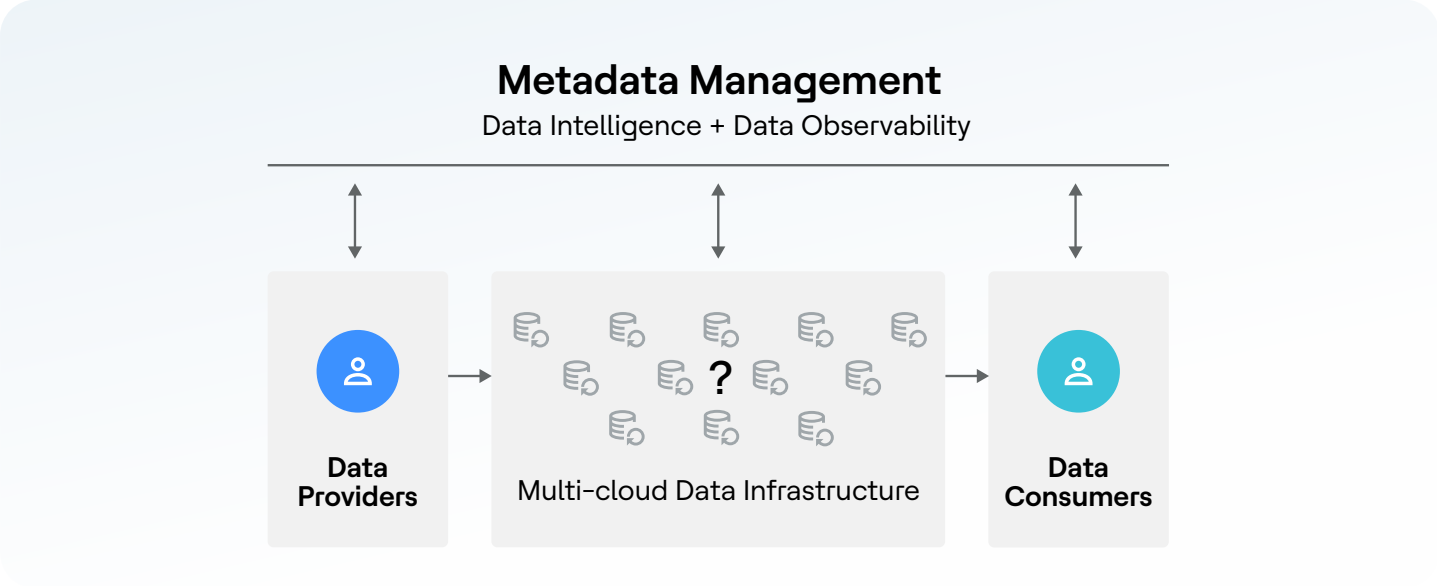
Ole Olesen-Bagneux is a globally recognized thought leader in metadata management and enterprise data architecture. As VP, Chief Evangelist at Actian, he drives industry awareness and adoption of modern approaches to data intelligence, drawing on his extensive expertise in data management, metadata, data catalogs, and decentralized architectures. An accomplished author, Ole has written The Enterprise Data Catalog (O'Reilly, 2023). He is currently working on Fundamentals of Metadata Management (O'Reilly, 2025). With a PhD in Library and Information Science from the University of Copenhagen, his unique perspective bridges traditional information science with modern data management.

ACTIAN™
a division of HCLSoftware

Why do Organizations Need Data Observability?

The answer is straightforward: Companies have immense amounts of data stored and transformed across many different technologies. To ensure that data is useful for the intended purposes, companies need to observe their data meticulously. Data observability has emerged to do exactly this, as a discipline and as a technology. [1]

Accordingly, some employees are tasked with providing data that they need to transport and transform, to make the data consumable for other employees, data consumers. To bridge providers and consumers of data, organizations over time build a large, multi-cloud data infrastructure – and this is the drama: This data infrastructure is difficult to control, as it consists of a plethora of technologies from multiple tech vendors that holds enormous amounts of data that is transported through many solutions in many ways. This is complicated to govern. To solve this, companies rely on metadata management to govern their data infrastructure. [2]



Think of the system owners and the data these systems contain in your organization: It can be data from domains such as HR, Manufacturing, R&D, Finance, etc. The data contained in these systems holds tremendous value–that data is the source of innovative use cases that can provide improved and even new offerings, ensuring growth and a prosperous future of your company.

The metadata layer on top of the multi-cloud data infrastructure provides companies with the intelligence to govern their data through a set of metadata capabilities, one of them being data observability.

In short, organizations need data observability because so much data is transported through so many storage solutions that manually controlling all this data is impossible. Data observability ensures fresh, correct and high-quality data for data consumers.

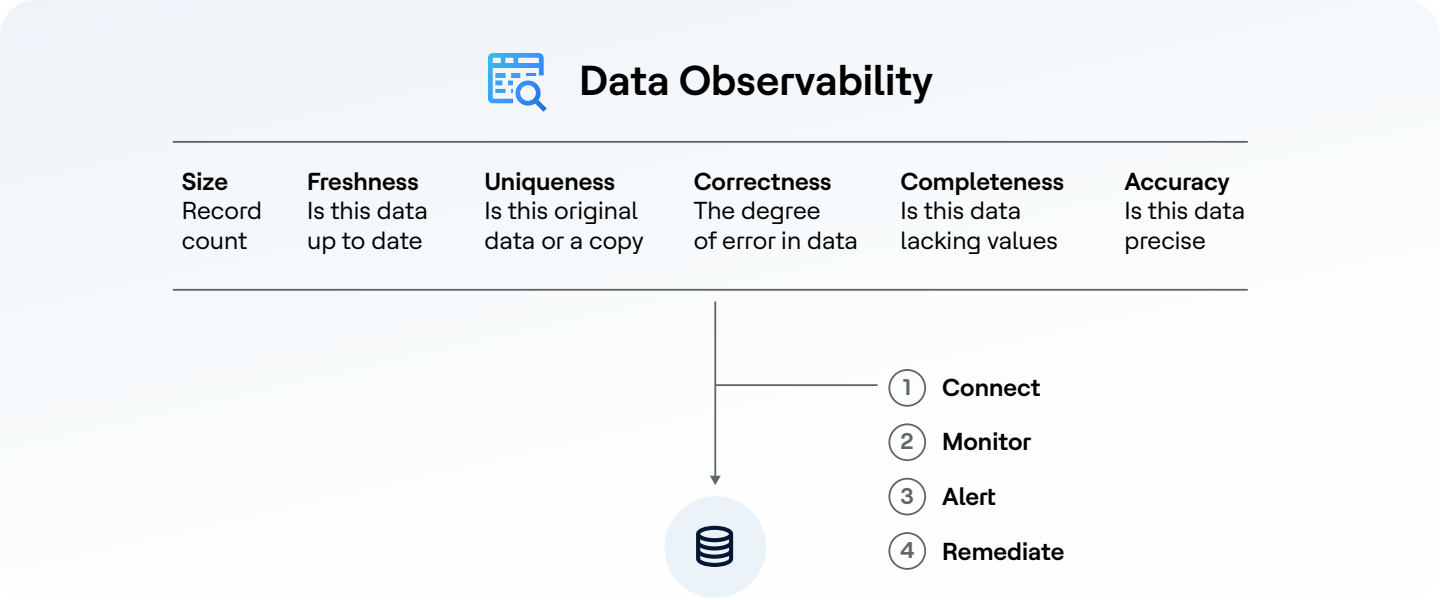
ACTIAN™ a division of HCLSoftware

What is Data Observability?

Let's dive a little deeper–what exactly is data observability? **Data observability is observing data in terms of its size, freshness, uniqueness, correctness, completeness, and accuracy. These metrics [3] can be set up, managed, and enforced by connecting, monitoring, alerting, and ultimately remediating data.**

Let's go through these in a little more detail.

Transporting data in pipelines. Transporting data through pipelines known as ETL (extract-transform-load) or ELT (extract-load-transform) tools is the core activity of data engineering. [4] To support data engineers, data observability can assess the performance of pipelines in terms of scope and stability. Observability allows reducing the cost of the pipeline with fewer, simpler steps and even warns if planned pipelines will be brittle or fail.



Data Observability enables:

- **Logical data pipelines (transportation of data):**
  - Data pipeline performance
  - Data pipeline cost
  - Data pipeline predictions
- **Scaling data architectures:**
  - Reliable data products
  - Easier implementation of data contracts
  - Increased success for AI use cases
- **Governance and Compliance**
  - Increase data quality
  - Identify PII and sensitive data usage
  - Enhance the data catalog (the "search engine") in the metadata layer

Scaling data architectures. Organizations are moving towards data products because data managed as a product is more scalable than data managed in big enterprise data warehouses and data lakes. Data observability is a key component to scale data architectures with solid pipelines for high-quality data products, made easy to consume through automated data contracts, ultimately making your company succeed with AI. [5]

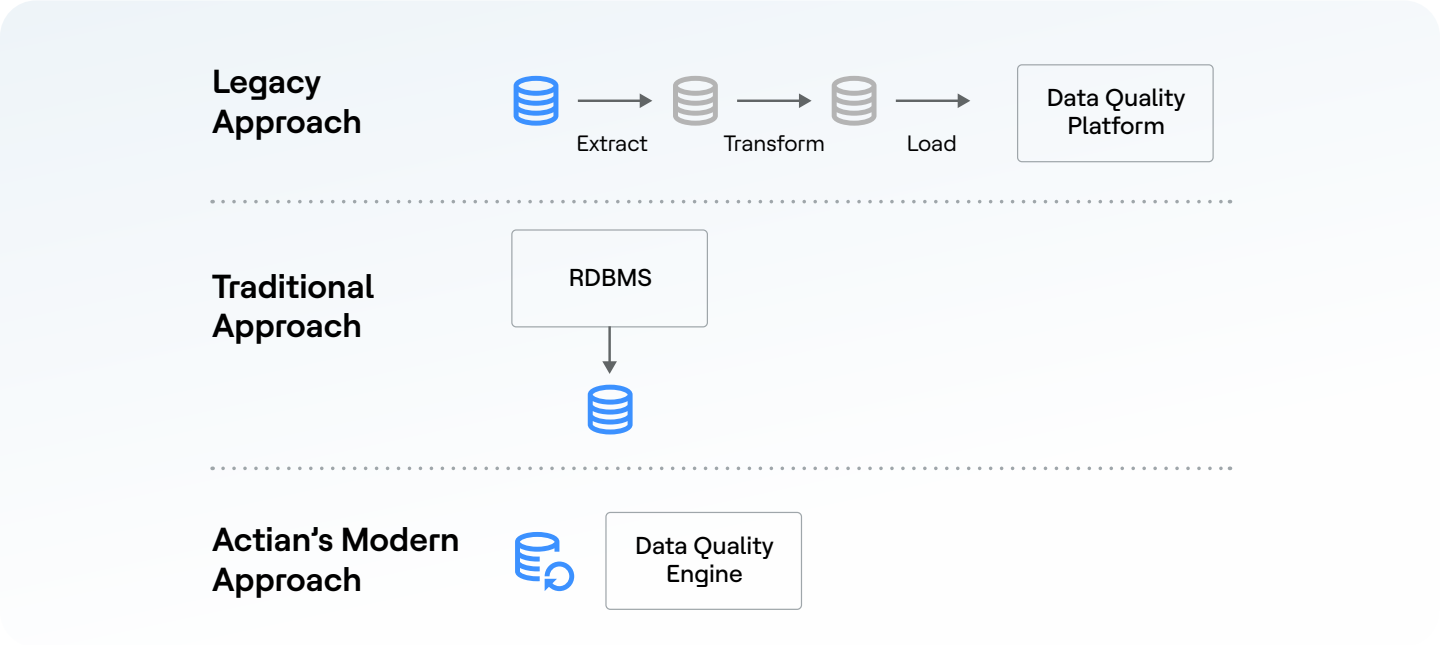
Governance and Compliance. Data observability plays a vital role in data governance, as it ensures data quality through policies that can be enforced with monitoring, including PII and sensitive data usage. Observability can even be built into the "search engine" for the metadata layer – the data catalog, so end users benefit from only discovering trusted data. [6]

ACTIAN™
a division of HCLSoftware

How is Data Observability Performed With Actian?

At Actian, we are championing a completely new and modern approach to data observability. Data observability is most often performed using a traditional approach, and some organizations still run it in a legacy setup. You can see the various approaches in the diagram below.

The traditional approach. The traditional data observability approaches use the Relational Database Management System (RDBMS) on top of the database to query and observe the data, just as every other type of query performed with the RDBMS. This approach has advantages compared to the legacy approach but still has substantial drawbacks. There is a computational limit to the number of queries you can run, and to be cost-effective, the queries can't run on all the data.



Let's explore what makes Actian's solution stand out by explaining the legacy, traditional, and modern approaches to data observability.

The legacy approach. The legacy approach to data observability involves setting up a classic data pipeline that extracts, transforms, and loads data into a data quality platform, where a combination of automated and manual checks is performed on the data. This setup is very time-consuming, not synchronized with the data it observes (changes can occur that make findings in the observed data obsolete), and it does not scale because it is too labor-intensive.

Therefore, the data is observed incompletely, because it can only perform its observations on sampling, making your data observability fundamentally unreliable. It doesn't scale to a level allowing you to observe all your data in each source. Also, there are substantial limitations in the types of data sources you can observe, leaving out semi-structured data, such as nested JSON scripts and YAML files. Finally, this approach is query-intensive, and that makes it expensive. And do you really want unreliable, incomplete, and costly data observability?

ACTIAN™
a division of HCLSoftware

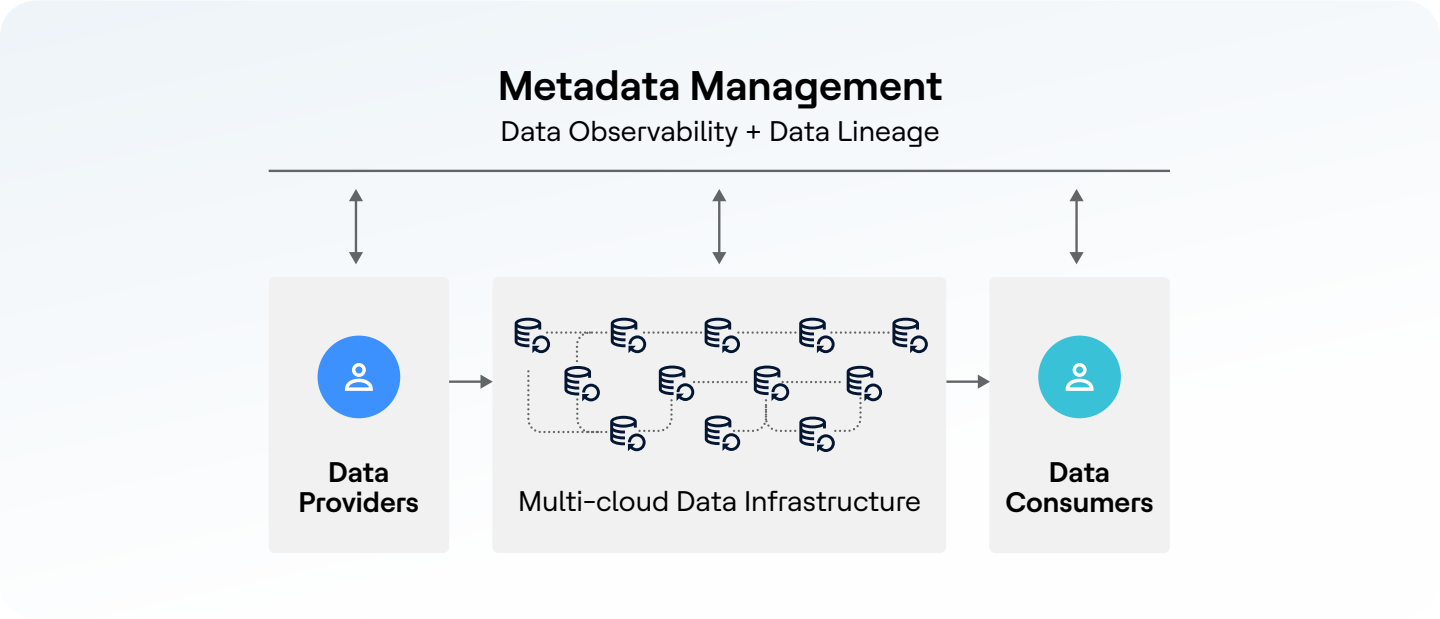
Actian's Modern Approach

To observe data, Actian Data Observability sets up a small, decoupled data quality engine alongside the data. This engine monitors data continuously in memory, next to data, not through the relational database management system above it. [7] This way, data is not transported to be observed, like in the legacy approach. Nor is the database queried with SQL to observe the data, like in the traditional approach. Because of using a modern approach with a decoupled data quality engine, Actian Data Observability can connect to all your structured and semi-structured data sources, from JSON, Parquet, YAML files to most open table formats and streams natively – which are typically unobservable with the traditional approach. One of our biggest advantages is how Actian Data Observability makes data quality completely independent from the underlying systems. This system-agnostic approach significantly simplifies adoption for customers with diverse tech stacks. Plus, this decoupled architecture future-proofs their quality processes by eliminating migration headaches down the road when systems change. Furthermore, since the data quality engine operates next to data, it takes pressure off operational systems by not running queries against their databases continuously.

This also significantly reduces the cost of data observability because of condensed, precise monitoring of memory. Finally, this approach can perform complete observability of all the data in the sources since it does not have to consider the cost of querying data. This provides you with complete data observability, which is easy to set up and run, it's cost-effective, and first and foremost, it is complete. One final key component makes Actian Data Observability powerful:

The Complete Solution - Data Lineage for Data Observability

By itself, a data observability solution cannot track data movement across data sources. It only looks directly at one source or more without comparing them. Thus, it cannot track the change in data quality as data moves across the data infrastructure unless you map sources manually. That is why Actian Data Observability is equipped with data lineage going forward, which automatically tracks data across data sources. This combination makes it possible to provide perfect observability from the data provider to the data consumer.



ACTIAN™
a division of HCLSoftware

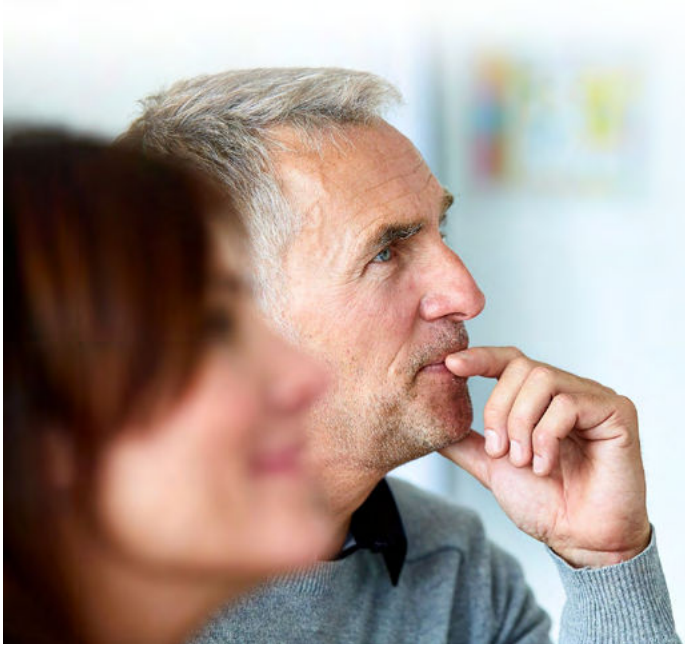
Executive Summary: Actian's Point of View on Data Observability

In Actian, we believe that data observability can be performed according to the technological values that we hold dear: data observability must be trusted, flexible and easy to use. That's what we deliver with Actian Data Observability.

As a data leader or data technology decision maker, you need to know this: Your organization deserves a modern, cost-effective, and complete data observability solution. Actian Data Observability stands out as exactly that in a market dominated by technologies that are, unfortunately, expensive, complicated, and quite simply incomplete.

You cannot rely on traditional data observability solutions to deliver on the strategic data projects that your company's future depends on. Traditional data observability technologies will not scale and will most likely not even get impactful adoption within your organization. However, they will stress your source systems and be expensive.

With Actian Data Observability, combined with Actian's data lineage capabilities, your company can build data confidence across your entire IT landscape while keeping production systems running smoothly and with minimum costs.



References

[1] We will discuss data observability relying on the definition, that we agree in, provided by Gartner®, in "Reference Architecture Brief: Data Observability," Prasad Pore, 10 September 2024: "Data observability is the ability to holistically understand the state and health of an organization's data, data pipelines, data landscapes, data infrastructures and the financial governance of the data. This is accomplished by continuously monitoring, tracking, alerting, analyzing and troubleshooting problems to reduce and prevent data errors or downtime."

[2] The diagram is inspired by Figure 1-1. in Piethein Strengholt: Building Medallion Architectures (O'Reilly, 2025) – data observability is a needed capability in medallion architectures too.

[3] We will not dive into the mathematical details of the metrics here, for a valuable source on the topic, see Andy Petrella: Fundamentals of Data Observability, (O'Reilly, 2023), chapter 4

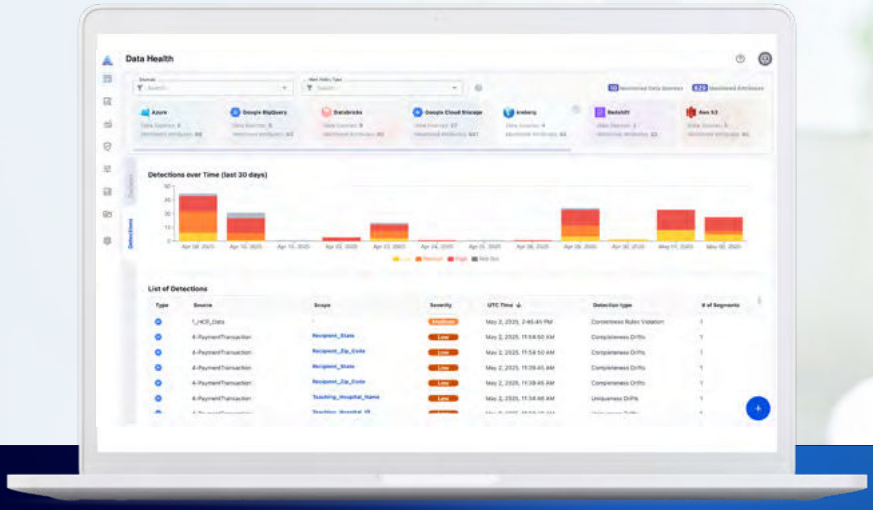
[4] The seminal work about data engineering is Joe Reis and Matthew Housley: Fundamentals of Data Engineering (O'Reilly, 2022)

[5] Studies show that over 60% of AI projects will fail before 2026 due to data not being AI ready, see: Gartner®" A Journey Guide to Delivering AI Success through "AI-Ready" Data" by Ehtisham Zaidi, Roxane Edilali, 18 October 2024

[6] Ole Olesen-Bagneux: The Enterprise Data Catalog, (O'Reilly, 2023), chapter 7

[7] Technical details about this setup are important, but too dense for this white paper. For further details see **actian.com/data-observability**

ACTIAN™
a division of HCLSoftware



Want to See Actian Data Observability in Action?

Start reducing costs, increasing trust, and simplifying governance across your entire data architecture—faster than you thought possible.

With Actian's modern, decoupled approach, you get full visibility into the quality of your data—without taxing your systems or your budget. Whether you're scaling data products, preparing for AI, or enforcing governance at speed, Actian helps you take control with confidence.

> Visit **actian.com** and **request a demo** to see how Actian can transform the way you observe, trust, and manage your data.

About Actian

Actian empowers enterprises to confidently manage and govern data at scale. Organizations trust Actian's data management and data intelligence solutions to streamline complex data environments and accelerate the delivery of AI-ready data. Designed to be flexible, Actian solutions integrate seamlessly and perform reliably across on-premises, cloud and hybrid environments. Learn more about Actian, the data division of HCLSoftware, at **actian.com**.

Tel +1 512 231 6000 Fax +1.512.231.6010 710 Hesters Crossing Road, Suite 250, Round Rock, TX 78681 **Actian.com**

© 2025 Actian Corporation. Actian is a trademark of Actian Corporation and its subsidiaries. All other trademarks, trade names, service marks, and logos referenced herein belong to their respective companies.

All Rights Reserved. V1-2025-5

ACTIAN™
a division of HCLSoftware