

# Data Observability Demystified: From Basics to Advanced

Turn observability into a business dependency. Catch issues early and fix them before bad data drives bad decisions.

## Table of Contents

- 4 Executive Summary: Data Health Is a Non-Negotiable
- 5 Understand the Basics: What is Data Observability?
- 5 The Business Imperative: Ensuring Trust in Data
- 6 The Human Side: Who Uses Data Observability and Why
- 7 The Technical Role of Data Observability
- 7 4 Business Use Cases: Where Observability Shows Up in Industries
- 9 4 Technical Use Cases: Supporting the Tech Side of Observability
- 10 7 Data Observability Agents and What They Do
- 11 MCP and Agentic Connectivity: Bringing Observability to AI Assistants
- 11 6 Steps to Modernize Your Approach to Data Observability
- 12 About Actian

## Executive Summary: Data Health Is a Non-Negotiable

Business leaders have become experts at measuring performance. They have dashboards everywhere and metrics for everything. The problem is that too many of those dashboards are like a speedometer with no fuel gauge. You can see how fast you're going, right up until the moment you sputter out.

Data observability is the fuel gauge and engine diagnostics that businesses need, but often lack. Observability ensures the data driving AI, automation, and decision-making is reliable and current. Without freshness checks, for example, forecasts can be delayed or based on outdated numbers, creating avoidable risk in everyday decisions.

Now with agentic AI in the mix, the data quality mission expands again. It's not enough to detect issues after the fact. You need systems that can validate data as it arrives, explain how it transforms over time in plain business language, and mitigate problems before unreliable data triggers unreliable outcomes.

That's why data observability is evolving with data observability agents. These agents accelerate root cause identification and keep reliable data flowing to the analysts, domains, and AI systems that depend on it.

## Understand the Basics: What is Data Observability?

Data observability is the ability to understand the health and state of the data in your systems by measuring key signals like quality, reliability, and lineage. This helps you answer three important questions quickly:

- Is the data wrong?
- What broke and where?
- How do we fix it?

A true observability approach goes beyond a single dashboard or point solution. It offers a holistic and automated way to monitor data, detect issues early, and shorten the path from incident to resolution across pipelines, tables, and critical business processes.

The stakes for quality data are rising. Organizations are investing heavily in AI and analytics, but projects fail when data quality and controls don't hold up and aren't trusted. A lack of a holistic view into your data health results in an incomplete narrative, limiting your insights.

Data must be reliable for business outcomes, including AI use cases, to also be reliable.

## The Business Imperative: Ensuring Trust in Data

Data observability isn't about firefighting data issues. It's about supporting faster decisions, better customer experiences, and stronger governance because teams can find and activate data they trust.

Quality data is essential for AI and other use cases. Gartner estimates that through 2025, at least 30% of AI projects will have been abandoned after proof of concept due to poor data quality, inadequate risk controls, escalating costs, or unclear business value. Gartner also notes that 93% of teams fail to detect data issues before they impact the business.

Data observability supports practical use cases such as:

- **Operational continuity.** Experience fewer disruptions if upstream systems change.
- **Customer experience protection.** Mitigate incorrect recommendations, pricing errors, or shipping issues.
- **Risk and compliance support.** Gain better visibility into sensitive data handling and lineage.
- **AI confidence.** Reduce overhead for automated workflows and trust the outcomes.

The cost of not catching quality issues early escalates as data moves downstream. For example, a small upstream schema change can cascade quickly (Figure 1).

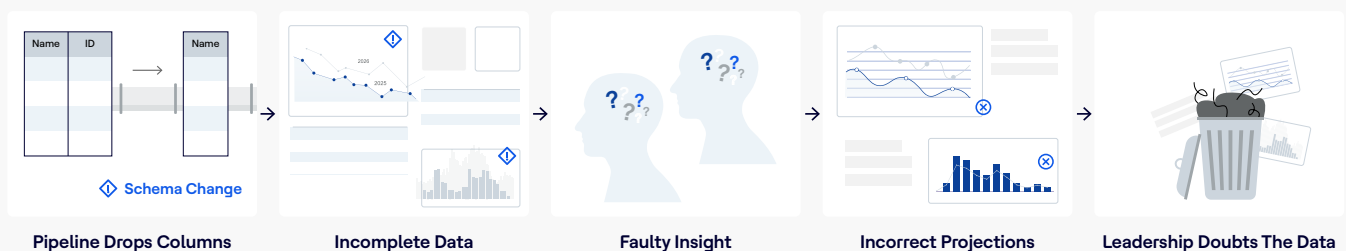


Figure 1.

A pipeline silently drops columns in a table, the dashboard experiences drift without complete information, decisions are based on the unreliable dashboard, forecasts are missed, and the leadership loses trust in the data.

Data observability breaks this chain of events early by surfacing the issue where it starts, before it impacts downstream use cases.

## The Human Side: Who Uses Data Observability and Why

Data observability only works if it solves real problems. Typical personas that use observability and the pain points it solves include:

|                         |                                                                                                                                                                                                                                                                                         |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Data engineers</b>   | Responsible for building and maintaining the data infrastructure and pipelines that feed observability systems. They need comprehensive pipeline monitoring to proactively identify and resolve issues before they impact downstream consumers and business processes.                  |
| <b>Data analysts</b>    | Specialize in processing and analyzing data, relying on high-quality, observable data for accurate reporting and insights. They need confidence that the data is accurate, trustworthy, and up-to-date, without questioning data quality.                                               |
| <b>Data scientists</b>  | Use advanced analytics techniques and depend on governed, high-quality data for model training and deployment. They need visibility into data quality and lineage to trust the datasets feeding their models and understand when model performance degrades due to data issues.         |
| <b>DevOps engineers</b> | Manage the technical infrastructure that supports data observability platforms and ensures system reliability. They need monitoring and alerting systems that help them maintain the infrastructure supporting data operations while ensuring minimal downtime and optimal performance. |

|                                              |                                                                                                                                                                                                                                                                           |
|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Business intelligence (BI) developers</b> | Create dashboards and reports that rely on observable, trustworthy data sources. They need observability tools that integrate with their reporting platforms to deliver accurate dashboards and alert business users when data quality issues might affect their reports. |
|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                            |                                                                                                                                                                                                                                                 |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Compliance officers</b> | Ensure that data management practices comply with regulatory requirements. They need comprehensive data lineage and audit trails down to the field level to demonstrate regulatory compliance and quickly respond to data governance inquiries. |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                    |                                                                                                                                                                                                                                                                                      |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>IT managers</b> | Oversee the technology infrastructure, which requires visibility into the performance and costs of data observability solutions. They need observability platforms that provide clear ROI metrics and help optimize data infrastructure costs while maintaining high service levels. |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                   |                                                                                                                                                                                                             |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Chief data officers (CDOs)</b> | Lead enterprise data strategy. They need comprehensive metrics on data quality, usage, and business impact to demonstrate the value of their data investments and identify areas for strategic improvement. |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Data trust should not live with only one department. It must be decentralized so business users can ask and answer observability questions without waiting in a queue. Natural-language interfaces and specialized reliability agents help by reducing bottlenecks and shortening time-to-resolution.



## The Technical Role of Data Observability

Observability helps ensure data is accurate and fit for use across systems and processes. It refers to monitoring, managing, and maintaining data in a way that ensures quality, availability, and reliability.

Achieving the business imperatives of data observability relies heavily on four key technical capabilities:

1. **Data monitoring.** See what's happening now. This is critical for applications that require immediate action, such as fraud detection systems and dynamic pricing strategies.
2. **Automated detection and resolution.** Spot anomalies and respond quickly. Modern data observability solutions must handle large datasets and support complex anomaly detection efficiently. This is vital in retail, finance, and telecommunications where the ability to detect and resolve issues automatically is needed to maintain operational efficiency and customer satisfaction.
3. **Data quality and accessibility.** Ensure data is usable and trusted. The quality of insights directly correlates with the quality of data ingested and stored throughout your systems. Data that's accurate, clean, and easily accessible is essential for effective analysis, reporting, and training AI models.
4. **Advanced capabilities.** Use AI-powered anomaly detection and automated root cause analysis. Data observability platforms are evolving to meet new challenges and opportunities by leveraging machine learning to identify unusual patterns.

A data observability solution enables complete visibility into data flows and dependencies while supporting rapid troubleshooting and issue resolution. All of this work is performed directly within the observability environment.

## 4 Business Use Cases: Where Observability Shows Up in Industries

Data observability directly supports critical business objectives across all industries. Sectors use observability to drive measurable outcomes:

### 1. Financial services: Identify risk and fraud

Enhance risk assessment and fraud prevention by detecting anomalies in transaction patterns. Also validate freshness for market feeds and support regulatory compliance.

#### Examples

-  **Banking.** Detect anomalies in transaction flows to identify potential fraud, while ensuring compliance with anti-money laundering regulations through comprehensive data lineage tracking.
-  **Insurance.** Monitor claims data and processing times to identify fraudulent patterns and ensure accurate risk assessment models.
-  **Investment management.** Track data freshness and quality to ensure portfolio management decisions are based on accurate, timely information.

### 2. Healthcare: Patient safety and operational excellence

When electronic health record (EHR) or admissions data is incomplete, delayed, or inconsistent, the downstream effects can be serious. Observability supports proactive monitoring of patient data quality, medical device outputs, and clinical workflow efficiency to improve patient outcomes and operational performance.

#### Examples

-  **Hospitals.** Monitor EHR data quality to ensure accurate patient information, medication dosing, and treatment recommendations.
-  **Pharmaceutical.** Track clinical trial data integrity and regulatory compliance throughout drug development processes.
-  **Medical devices.** Ensure IoT sensor data from patient monitoring equipment maintains accuracy and availability for critical care decisions.

### 3. Retail and eCommerce: Customer experience

Product catalogs, inventory availability, recommendations, and pricing all depend on trusted pipelines. Observability helps prevent silent errors that lead to customer frustration. It's the business equivalent of an air traffic controller for your business system, continuously tracking movement, detecting anomalies early, and intervening before small issues become visible failures.

#### Examples

-  **eCommerce platforms.** Monitor product catalog data quality, inventory accuracy, and recommendation algorithm performance to boost customer satisfaction and prevent lost sales.
-  **Omnichannel retail.** Ensure customer data consistency across online, mobile, and in-store touchpoints for seamless shopping experiences.
-  **Supply chain.** Track inventory data quality and supplier performance metrics to optimize stock levels and prevent out-of-stock or overstock situations.



**Consumer goods.** Analyze production data patterns to optimize batch processing and reduce waste while maintaining quality standards.

Continuously monitoring pipelines for freshness, accuracy, and unexpected change lets you trust what you see, act faster, and avoid the hidden costs of bad data before it affects the bottom line.

## 4 Technical Use Cases: Supporting the Tech Side of Observability

Ever wonder how business strategies become digital reality? These use cases pull back the curtain on the technical muscle behind data observability. They show how observability turns business intent into reliable data, actionable insights, and competitive advantage:

### 1. Data pipeline monitoring and automated recovery

Continuously monitor complex ETL/ELT pipelines and automatically recover from certain types of failures. This ensures the resilience required for mission-critical data operations.

#### Key capabilities

- **Pipeline health monitoring** with automated alerting for job failures, performance degradation, and data quality issues.
- **Automated recovery mechanisms** for common pipeline failures, including retry logic and fallback data sources.
- **Performance optimization recommendations** based on historical execution patterns and resource utilization.

### 4. Manufacturing: Production optimization and quality control

Gain insights into production data quality, equipment performance, and supply chain efficiency to optimize manufacturing processes and ensure products meet the proper specifications.

#### Examples

-  **Automotive.** Monitor sensor data from production lines to ensure quality control standards and predict equipment maintenance needs.
-  **Aerospace.** Track component traceability data and quality metrics throughout the manufacturing process to ensure safety and regulatory compliance.

## 2. Data quality scoring and SLA management

Implement automated data quality scoring to measure data health against defined SLAs and business requirements. This way, “good enough” data is measurable, enforceable, and visible.

### Key capabilities

- **Automated data quality scoring** uses configurable weights for different quality dimensions.
- **SLA tracking and reporting** with dashboards showing performance against established metrics.
- **Proactive alerting** when quality scores fall below acceptable thresholds.
- **Historical trend analysis** to spot patterns in degradation.

## 3. Incident response and root cause analysis

Detect incidents, speed triage, and support root cause analysis. Reduce mean time to resolution (MTTR) when data issues arise.

### Key capabilities

- **Intelligent incident correlation** that groups related issues and surfaces likely root causes.
- **Automated impact analysis** showing which downstream systems and users are affected by issues.
- **Collaborative incident management** with integrated communication tools and runbook automation.
- **Post-incident analysis** with automated report generation and recommendations to prevent repeats.

## 4. Data lineage and impact analysis

Provide end-to-end data lineage and impact analysis to understand dependencies, manage change safely, and connect technical assets to business meaning.

### Key capabilities

- **Automated lineage discovery** across complex, multi-vendor data ecosystems.
- **Column-level lineage tracking** showing detailed transformations and dependencies at the field level.
- **Change impact analysis** predicting downstream systems affected by upstream changes.
- **Business glossary integration** connecting technical lineage to business terms and definitions.

Organizations used to rely on observability solutions to tell them when a data pipeline was broken. The new expectation is for the solution to detect quality issues early, explain them clearly, and help fix them quickly.

### Take the Next Step: Data Observability for Modern Workflows

Data observability agents take a modern approach to quality by validating data at the source rather than downstream. Traditional data quality strategies often catch problems after data has already moved through pipelines and reached data consumers.

Moving the process upstream, where data is first made available for use, keeps poor quality data from being used by dashboards, applications, or AI workflows.

## 7 Data Observability Agents and What They Do

As expectations shift from data issue visibility to prevention, observability needs more capabilities. That's where data observability agents come into play.

Purpose-built data observability agents work together across the lifecycle from data detection to resolution to deliver and maintain quality data. Actian offers these seven data observability agents with specialized capabilities:

1. **Validation Agent.** Recommends and applies validation rules such as schema checks, completeness, consistency, and freshness, at ingestion. Prevents "bad data" from entering downstream paths.
2. **Incident Diagnosis Agent.** Detects anomalies and explains likely root causes in plain language, like upstream feed change, job failure, or sudden null spike.
3. **Lineage Agent.** Maps dependencies so you can see a "blast radius" of any tables, dashboards, and models impacted by a failing data asset.
4. **Data Insight Agent.** Translates quality signals into stakeholder-friendly guidance, such as what changed and what it affects. Builds trust in the data with easy-to-understand insights.
5. **Orchestration Agent.** Automates resolution steps and integrates with ticketing and alerting workflows to close the loop from detection through remediation.
6. **Routing Agent.** Notifies the right owners based on domain, severity, and impact. Prevents issues from getting stuck in channels that don't have direct ownership.
7. **Help Agent.** Guides users through setup and troubleshooting with natural language, which helps create monitors, understand incidents, and improve quality over time.

Together, these agents move observability from reactive monitoring to proactive, autonomous tasks for reliability you can trust.

## MCP and Agentic Connectivity: Bringing Observability to AI Assistants

Data teams will increasingly work with AI assistants and agents. A model context protocol (MCP) compliant server allows AI agents to access observability context without custom integrations.

An MCP layer ultimately lets agents retrieve validated data along with the quality metadata required for informed decisions. Think of MCP as the "USB-C" for AI-to-tool connectivity. It's a standard interface that reduces one-off connections and accelerates safe adoption across AI assistants and agent frameworks.

### Why Actian Data Observability Is Different

Most observability tools are optimized for one primary platform and struggle when data is spread across data lakehouse components, open table formats, and multiple clouds. Actian takes a different approach by offering data observability that's designed for breadth, openness, and in-place reliability across the lakehouse ecosystem.

Actian differentiators include:

- **Microsoft Fabric / OneLake coverage.** Monitor data in place. Profile and monitor directly at the OneLake layer so Microsoft Fabric users can benefit from continuous reliability without moving data.
- **Catalog-agnostic observability for open lakehouse stacks.** Actian gives you broader coverage regardless of which catalog you standardize on.
- **Designed for federated architectures, not single-platform assumptions.** Actian supports open, federated data architectures where data lives across multiple platforms and clouds, so reliability doesn't break if your architecture does.

If your lakehouse is multi-platform, multi-format, and constantly evolving, Actian helps ensure observability stays consistent. Trust doesn't depend on where the data happens to live.



## 6 Steps to Modernize Your Approach to Data Observability

If you're modernizing data observability for an AI future, focus on progress over perfection. These steps can help:

1. **Start with critical assets.** Prioritize data that's supporting revenue, risk, patient safety, or customer experience.
2. **Define trust signals that match outcomes.** Ensure data freshness, completeness, schema stability, and distribution stability that's mapped to actual SLAs.
3. **Automate detection and explanations early in the process.** Reduce the time to understand issues and use it as an ROI lever.
4. **Close the loop with orchestration.** Detection without action does not advance quality. Integrate routing, ticketing, and remediation workflows.
5. **Build trust for analysts.** Decentralize trust so business users can confidently interpret data health.
6. **Make AI agents trust-aware via MCP.** Let agentic workflows check quality context before acting.

Data observability started as visibility. It's now evolving into something more operational, which is data reliability that supports automation and agentic decision-making without sacrificing trust.

Data observability agents are built for that shift by combining proactive ingestion validation, autonomous detection and remediation, plain-language explanations, and MCP-connected workflows. This allows both humans and AI agents to operate with the same trusted context.

If your organization is serious about agentic AI, the question isn't whether you need observability. It's whether your observability is ready to support autonomy with guardrails.

## About Actian

Actian empowers enterprises to confidently manage and govern data at scale. Organizations trust Actian 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 and AI division of **HCLSoftware**, at [actian.com](https://actian.com).

