Real Talk: A Guide To Understanding Data Quality & Data Observability

Navigating the ever-evolving maze of data quality management with definitions, head-to-head comparison, and insight from enterprise data teams.

Table of Contents

Introduction

The challenge of delivering quality data — accurate, reliable, complete, and fit for the purpose at hand — is nothing new for data teams. But the technology and terminology around how teams can improve data quality has been evolving at a breakneck pace.

If you find yourself confused by the current landscape, you're not alone.

For example: this year, Gartner published its first Market Guide for Data Observability Tools. In it, they report that "Vendors are offering a range of different capabilities branded as data observability, causing confusion in the market and tool adoption issues." Despite the lack of clarity, Gartner analysts predict that by 2026, 50% of enterprises implementing distributed data architectures will have adopted data observability tools, up from less than 20% in 2024.

But what is data observability? And how does it differ from traditional data quality solutions?

Let's find out.

The Ever-Evolving Maze of Data Quality Management

In our-ever evolving technology landscape (don't even get me started on GenAI…), the confusion is real.

But evaluating the feature set of every vendor to separate marketing buzz from reality is a tedious, time-consuming exercise. And translating capabilities into a clear understanding of how each type of product fits within your overall strategy, meets your requirements, and actually solves your data quality problems? That takes considerable research and expertise.

So we've done the work for you. Admittedly, as the category creator in data observability, we're a little biased. To keep us honest, we'll be referencing trusted third parties, highlighting the architectures of real data teams, and focusing on the core concepts that influence each solution category.

By the end of this guide, you'll have a concrete understanding of:

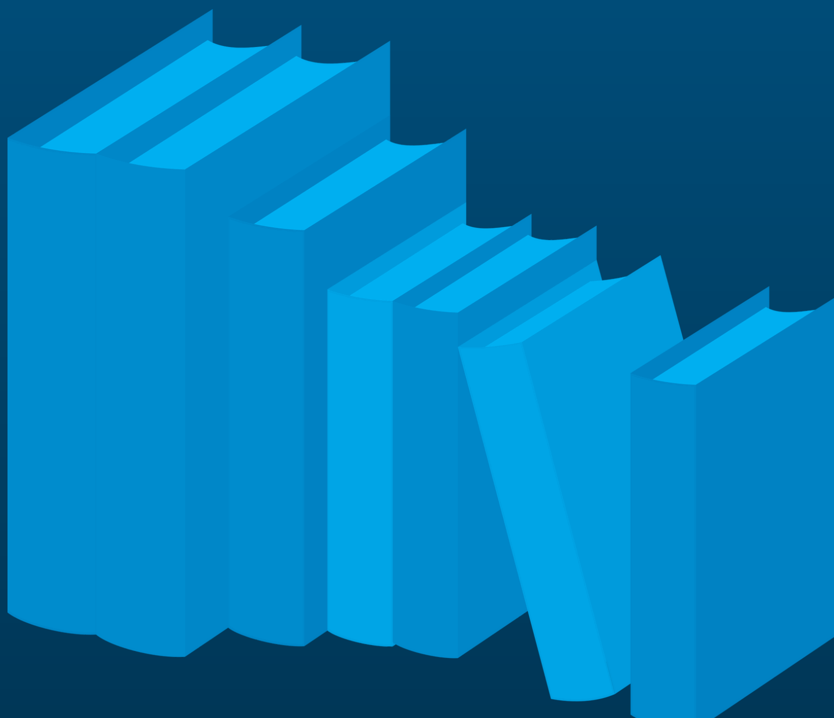
1. The purposes of and differences between data quality and data observability
2. The core concepts around data quality management: detection, resolution, and management
3. How to develop requirements that align with your team's strategic objectives, rather than focusing on individual technologies
4. The role of data observability in supporting AI-ready data and other emerging technologies

With the right context and a clear understanding of how these technologies each work towards improving data quality, you can navigate the maze of data quality management with confidence.

Let's start with a few definitions — demystifying these core concepts and empowering you to make informed decisions for your organization's data strategy.

Defining Data Quality, Data Reliability, & Data Observability — & Why They Matter

Data quality, data reliability, data observability — to the layperson, these terms sound nearly interchangeable. But in the context of data quality management, we have to dig deeper and understand some subtle but important distinctions.



Data Quality: A Snapshot in Time

Data quality is the extent to which data is fit for use, and it refers to the state of your data at a given moment. It's usually assessed across six key dimensions:

1. Accuracy—how well does your data reflect reality? Accuracy ensures that the information in your databases is error-free and aligns with the real-world entities it represents.
2. Completeness—is all the necessary data present? Completeness reflects whether any data is missing, and if your data has the sufficient breadth and depth for the intended purpose.
3. Consistency—do copies of the same data hold the same value across different systems or databases? Consistency describes the degree to which your data remains uniform across your entire data ecosystem, and the extent to which related data is in alignment around dimensions like definition, value, range, type, and format (as applicable).
4. Uniqueness—are there duplicates in your data? Uniqueness helps identify and eliminate redundant information.
5. Timeliness—was the data updated on time? Timeliness (also referred to as freshness) ensures that your data is current and reflects the most recent information available.
6. Validity—does your data fit logical conditions? Validity ensures that your data conforms to specific formats, falls within acceptable ranges, and adheres to business rules or logical constraints. For example, in a column for car models, a value of "Prius" would be valid, while "Monkey" would not.

One key concept here: data quality is measured as a snapshot in time. It captures the state of your data's health at a specific moment, but doesn't reveal any information about the broader data ecosystem.

Data Reliability: Fit for Purpose Over Time

If data quality is the weather, data reliability is the climate — it's the longer view, assessing patterns and changes in data quality over time. Data reliability asks: Is your data fit for purpose over time and across various conditions? When issues arise, how quickly can your teams recover?

For example, if you measure a product's data quality, that could mean assessing its availability at 9am, the completeness of its records, and its consistency versus a source-of-record — reflecting a snapshot of its data quality at a given moment. But in order for that data product to demonstrate data reliability, you have to assess whether it maintains those service levels over time, across holiday traffic spikes and splashy product launches.

Inspired by the discipline of Site Reliability Engineering (SRE), data reliability reflects an organization's ability to deliver high data availability and health throughout the entire data lifecycle.

Measuring data reliability is a function of measuring data quality attributes over time. If you set a data freshness Service Level Agreement (SLA) of one hour for a critical data product, you can assess its data reliability by how often you meet this SLA (say, 99% or 95% of the time).

Learn more about data reliability.

Data Observability: Full Visibility Into Data Reliability

Data observability goes one step further. It doesn't just ask if the data is reliable — it helps you understand why it might not be by identifying what went wrong, where, and when, at any point in the data lifecycle.

This requires having full visibility into your data, systems, and code so that you can ensure data reliability. Data observability tools don't just capture data quality metrics of your datasets, but monitor end-to-end data reliability and enable incident resolution across the entire data ecosystem.

As Gartner officially defines the category of data observability tools:

"Data observability tools are software applications that enable organizations to understand the state and health of their data, data pipelines, data landscapes, data infrastructures, and the financial operational cost of the data across distributed environments. This is accomplished by continuously monitoring, tracking, alerting, analyzing and troubleshooting data workflows to reduce problems and prevent data errors or system downtime."

In this way, data observability platforms provide insights into SLA adherence — how closely your data meets the requirements, based on specific use cases. These SLAs are ultimately driven by business needs and are crucial in ensuring that your data remains an asset, not a liability.

For instance, a retail company might require 99.9% uptime for its inventory data during peak shopping seasons. Data observability would not only monitor this uptime, but also provide insights into any factors that might threaten it.

The Chief Distinction Between Data Quality & Data Observability

The biggest difference to keep in mind: While both data quality solutions and data observability platforms monitor data quality, only data observability platforms enable organizations to improve it.

As Gartner points out, "Data quality is concerned with data itself from a business context, while data observability has additional interests and concerns the system and environment that deliver that data."

As your organization's data and AI use cases evolve, so too does the demands put on your organization to deliver highly reliable data. with multiple users, stakeholders, and use cases. Data quality solutions are a great first step when it comes to detecting issues, but they can't tell you what broke, who was impacted, and how to solve the problem.

Data observability takes your company to the next phase of the data quality maturity curve by providing visibility across the data, systems, and code powering your data estate, and providing a single pane of glass for engineers, analysts, governance managers, and executives to communicate and manage data reliability at scale.

We'll explore exactly why and how observability provides this actionable visibility in a moment — but first, let's take a step back and understand the big-picture concepts behind modern data quality management with end-to-end data observability.

The Fundamentals of Data Quality Management: Detection, Resolution, & Measurement

Regardless of which solution you choose to deploy, you need to understand the fundamental approaches that data teams can take.

We'll explore the core principles of detection, triage, resolution, and measurement, which form the backbone of any effective data quality strategy. Then, we'll compare how data quality and data observability tools approach each concept differently.

Detection

Detection is the first line of defense in maintaining data quality. This helps you identify when something's gone wrong with your data — ideally avoiding a 5am ping from your CEO, panicking because the dashboards are looking wonky right before the big investor meeting.

There are several methods you can use to detect potential issues:

1. Rules: comparison to a standard

This method involves setting predefined rules or thresholds for your data. For example, you might set a rule that all email addresses must contain an "@" symbol.

2. Anomaly detection: comparison to past behavior

By analyzing historical data patterns, you can identify when current data deviates significantly from the norm. This method is particularly effective for detecting sudden changes that might indicate issues.

3. Comparison to reality

This involves cross-referencing data points to ensure they align with real-world facts. For instance, if a data entry shows a zip code in one column and "Florida" in another, but the zip code isn't actually in Florida, this would flag as an inconsistency.

4. Comparison to other internal data

Similar to the reality check, this method compares data across different internal sources to ensure consistency. It's particularly useful for entity matching and maintaining data integrity across systems.

Learn more about data incident detection.

Triage

Once an issue is detected, the next step is triage — alerting the right team with the right context. But given the sheer number of data quality issues that can occur in a modern ecosystem, there's a real risk of drowning data teams with notifications.

To avoid alert overload and maximize efficiency, triage should include two key components:

Ownership

In complex data ecosystems with overlapping systems and teams, it's crucial to clearly define ownership — who is responsible for what. This ensures that when issues arise, they're quickly directed to the right team for resolution.

Severity and impact

Data incident alerts should include context around severity and potential business impact, so teams can prioritize responses accordingly.

Learn how [three real data teams approach triage strategies](#).

Resolution

Once data issues have been detected and alerted to the right owners, there are several approaches to resolving data issues:

1. Automatic remediation

Some systems can automatically fix certain types of issues. This sounds appealing, but proceed with caution — given the complexity and interdependence of modern data systems, remediation can potentially introduce more problems than it solves, and rarely addresses the root cause of recurring issues.

2. Prevention

Prevention involves implementing measures like circuit breakers and data contracts to prevent issues before they occur. It's a proactive approach that can significantly reduce the occurrence of data problems.

3. Manual troubleshooting

The system flags the problem for a person to investigate and resolve (ideally the right team member, with useful context, as we just described).

4. Root Cause Analysis (RCA)

The most advanced systems not only flag issues but also provide insights into the potential root cause, allowing team members to quickly identify and fix the underlying problem

Learn more about incident resolution.

Measurement

As the old saying goes, you can't manage what you can't measure. The same is true for data quality. To truly understand and improve your data reliability over time, your dashboards and scorecards should reflect:

1. Data quality

Track what rules or anomalies are currently raised, and ideally, the weight given to each based on potential impact. This provides a snapshot of your current data quality status.

2. Data reliability

Beyond just the quality of the data itself, measure the uptime of your data systems to understand how dependable your data infrastructure is over time.

3. Multiple units of measurements

To get a holistic view of your data health, measure data quality and reliability at multiple levels — tables, data products, and domains.

4. Time-to-detection

Measuring how quickly your team is detecting data quality issues is a helpful metric to understand the efficacy of your monitoring strategy and how new tooling is supporting this over time.

5. Time-to-resolution

Similar to the previous, time-to-resolution can tell you how well your tooling and processes are enabling your team to resolve data quality issues efficiently after detection.

6. Monitoring coverage

Similar to the previous, time-to-resolution can tell you how well your tooling and processes are enabling your team to resolve data quality issues efficiently after detection.

Learn more about incident resolution.

Augmented Data Quality Solutions

As the name suggests, augmented data quality solutions are focused on the quality of your data itself at a given point in time — not the underlying pipelines, platforms, and infrastructure, as well as the data in motion.

As Ventana Research analyst Matt Aslett describes, data quality tools are "concerned with the suitability of the data to a given task" and "designed to help users identify and resolve problems related to the validity of the data itself".

Understanding how data quality solutions were designed to solve specific problems is key to contextualizing their role in the current data stack — so let's take a brief look at the origins and evolution of these products.

Evolution & critical capabilities

The roots of augmented data quality solutions can be traced back to the early days of data management— a time of small-scale data operations and fewer interconnected systems, where manual checks and basic testing got the job done. But as data volumes grew and "big data" technologies like Hadoop emerged, companies began to offer more sophisticated data quality tools.

Today, Gartner defines data quality solutions this way:

"a set of capabilities for enhanced data quality experience aimed at improving insight discovery, next-best-action suggestions, and process automation by leveraging AI/machine learning (ML) features, graph analysis, and metadata analytics"

This emphasis on AI is the reason Gartner added "augmented" to the product category in 2024. AI-powered features are now considered essential to deliver critical data quality capabilities like:

- Active metadata support: Integration with data management tools to provide context and lineage information.
- Data transformations: The ability to modify data to fit predefined standards or business rules.
- Match, link, merge: Functionality to analyze and reconcile data across internal or external datasets, ensuring consistency and completeness.
- Profiling: Tools to identify characteristics or metrics of data within a table, providing insights into data structure and content.
- Rule discovery, management, and creation: Capabilities for setting, managing, and automatically discovering rules for data quality.
- Workflow and issue resolution: Features for managing the process of addressing and resolving data quality issues, including escalation and ticket management.
- Usability: Interfaces and features designed to support non-technical users, broadening the tool's accessibility within an organization.

In other words, we've come a long way from basic testing and manual checks. There have been significant advancements in automation and scaling — but still, it's important to understand these solutions remain focused on the quality of the data itself, not the surrounding pipelines and interconnected systems.

Future outlook — and limitations

If we look into our crystal ball at the future of augmented data quality solutions, both promising developments and potential challenges come to light:

Adoption of machine learning: Many solutions are incorporating machine learning techniques to enhance their capabilities, particularly in areas like anomaly detection and rule discovery. This trend is likely to continue, potentially improving the accuracy and efficiency of data quality processes.

Breadth vs. depth: While augmented data quality solutions offer a wide range of capabilities, there's a risk of these tools becoming "jacks of all trades, masters of none." As they spread their focus across multiple areas, it may become challenging for them to excel in any single aspect of data quality management.

Integration challenges: As data environments continue to grow more complex, these solutions may face challenges in seamlessly integrating with all components of a modern data stack.

Scalability concerns: With the exponential growth of data volumes and velocities, some augmented data quality solutions may struggle to scale effectively — particularly those built on cloud architectures.

Augmented data quality solutions have come a long way, and currently play an important role in many organizations. But consider their limitations and future trajectory when planning your data quality strategy.

Data Observability Platforms

Data observability platforms, on the other hand, have a much broader scope. These tools evaluate and provide visibility into data health across your data, systems, and code — not just the data itself.

As Ventana Research's Aslett describes, data observability platforms "...monitor not just the data in an individual environment for a specific purpose at a given point in time, but also the associated upstream and downstream data pipelines. In doing so, data observability software ensures that data is available and up to date, avoiding downtime caused by lost or inaccurate data due to schema changes, system failures or broken data pipelines."

Similar to data quality solutions, we can learn a lot about data observability by understanding its (much more recent) origin story.

Origination: cloud migration & the modern data stack

In the mid-2010s, data teams started migrating to the cloud and adopting new storage and compute technologies — like Redshift, Snowflake, Databricks, and GCP — that made data faster to process, easier to transform, and far more accessible.

But that also led to more complex pipelines and new personas (like data engineers) to manage the chaos. Companies gained tremendous value in terms of scalability and flexibility, but basics like data quality were often neglected and took a back seat. Products and BI tools were fed inaccurate data, and stakeholders would Slack data engineers at all hours, asking why their dashboards looked wrong.

The rise in what we call 'data downtime' directly led to the advent of data observability.

In 2019, our co-founder and CEO Barr Moses coined the term when the Monte Carlo platform first launched to help data teams leverage the well-established principles of application observability and performance monitoring (APM). And just as APM provides real-time insight into application performance and reliability, data observability aims to do the same for data systems.

Evolution: A comprehensive approach

Within a very short timeframe, data observability has evolved dramatically — from a nice-to-have to a 'must-have' solution for enterprise companies. Especially with the emergence of GenAI, leading data teams are urgently working to shore up data quality in order to fuel AI-powered products and applications. As a result, there are now nearly 50 listings in G2's Data Observability category, with hundreds of user reviews.

Currently, Gartner defines data observability as requiring several critical features, chiefly:

- Monitor and detect
- Alert and triage
- Investigate

They also note that some data observability solutions also include recommendations to resolve and prevent data issues.

Gartner clarifies these functions should occur across five key areas:

- Data: Monitoring the quality, structure, and content of the data itself.
- Pipelines: Observing the flow of data through various processing stages.
- Infrastructure: Monitoring the underlying systems that store and process data.
- Users: Understanding how data is accessed and used within the organization.
- Costs: Tracking and optimizing the financial aspects of data operations.

This is a good start. But at Monte Carlo, as the category creators and most established vendor in this space, we've found that an effective data observability approach needs to go beyond these categories to provide a truly comprehensive view.

Monte Carlo's approach to data observability

We think about our own evolution of data observability as focusing on four critical areas:

Anomalies to root cause: With end-to-end lineage, Monte Carlo helps data teams quickly investigate and resolve incidents by understanding all upstream and downstream relationships, from ingestion to analytics.

Data: This includes monitoring for anomalies in data freshness, volume, schema, and distribution. By analyzing historical patterns, we can automatically detect unexpected changes that might indicate data quality issues. Teams can also implement custom rules or profile their data within our platform to monitor the health of key assets exactly as their business needs demand.

Systems: We observe the health and performance of data infrastructure, including data warehouses, data lakes, ETL/ELT tools, and business intelligence platforms. This allows us to identify system-level issues that could impact data reliability.

Code: By monitoring changes in SQL queries, data models, and other code that interacts with data, we can pinpoint issues introduced by code changes or inconsistencies.

This approach allows us to not only detect anomalies but automatically correlate them to their root causes, spanning across data, systems, and code — and resolving incidents up to 80% faster.

MC MONTE CARLO

Future outlook

Data observability has already made a tremendous impact, but we believe it's just getting started. The future looks promising, with several key trends emerging:

Expanded coverage: As data ecosystems continue to evolve, we anticipate data observability platforms will expand to cover more types of systems and pipelines. This includes better integration with emerging technologies and data platforms.

Advanced analytics and AI: The use of machine learning and artificial intelligence in data observability will likely increase, enabling more accurate anomaly detection and predictive insights.

Focus on data products: As organizations increasingly treat data as a product, data observability will play a crucial role in ensuring the reliability and quality of these data products.

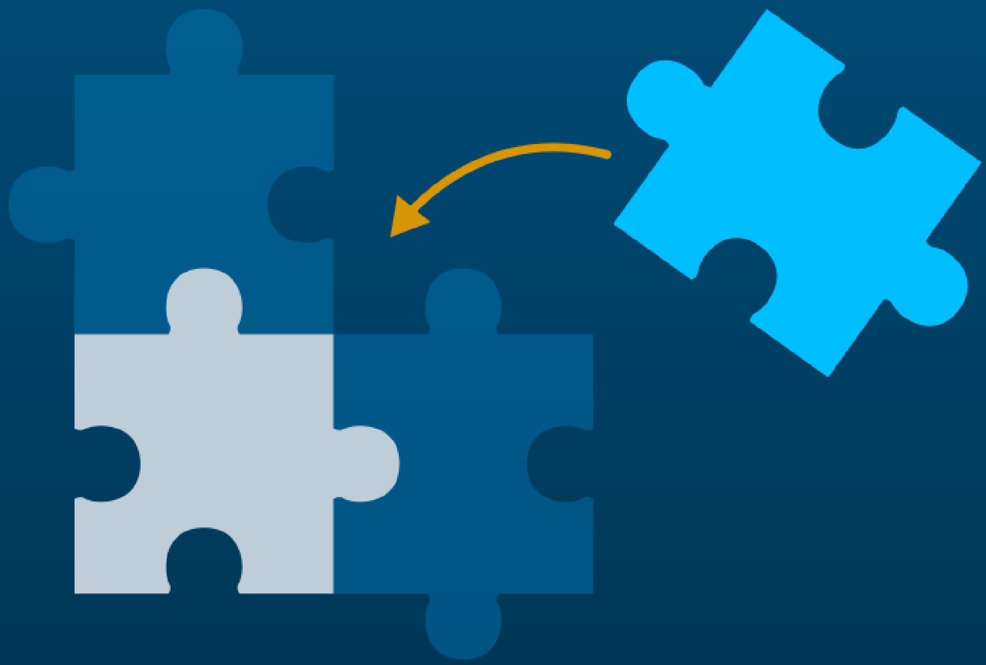
Support for AI/ML operations: With the growing importance of AI and machine learning, data observability will expand to support the specific needs of AI/ML models, including monitoring for data drift and model performance.

As GenAI continues to advance, data systems are only going to become more complex and crucial. And data observability promises to play a crucial role in keeping data reliable and trustworthy as the ecosystem evolves.

How Do Data Quality & Data Observability Compare?

Now, let's get down to pros and cons — or, in some cases, apples and oranges.

Here's how data quality solutions and data observability platforms compare in their respective capabilities to ensure data reliability.



Detection: AI-First vs. Rule-Based

There are two main differences between how data quality and data observability solutions approach detection: a) the type of monitors used; and b) how they're being deployed.

Traditional data quality solutions: rule-based and reactive

Data quality solutions typically rely on validating tables against predetermined rules, with some supplemental AI anomaly detection. Typically, these recommended rules are deployed after a table is scanned.

The downside? You can't possibly anticipate and create rules to capture all the ways your data can break, so there will inevitably be gaps. (We call these the "unknown unknowns" of data quality.) Plus, this manual process is an inefficient use of data engineering resources —both in the time spent and the compute power needed to hyperscale thousands or millions of rules.

Data observability: AI-driven and proactive

Data observability flips the script, using AI-first monitors that detect anomalies compared to historical patterns in your data, along with the ability to add custom rules for specific business cases. These AI monitors can be deployed across all tables, upstream of a data product as those tables are created. Data observability also offers AI-based rule suggestions via native profiling, making it easy to cover all of your data quality bases.

The result is more efficient, no-code monitor deployment, more comprehensive coverage, and — as we'll see shortly — better incident response.

Triage: Isolated Alerts vs. Integrated Lineage

The key difference in how data quality and data observability solutions approach triage lies in how alerts are managed and contextualized.

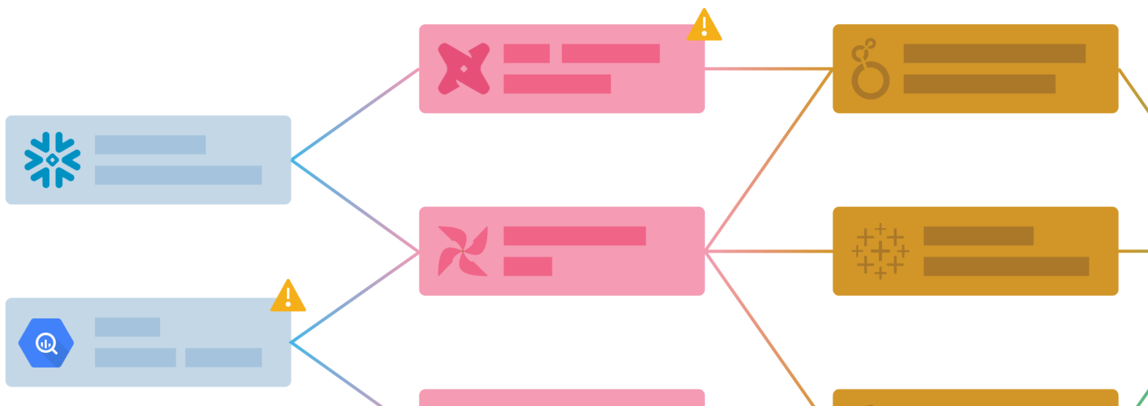
Traditional data quality solutions: isolated — yet overwhelming — alerts

In data quality solutions, each rule alerts independently — and the downstream impact is a mystery. One pipeline failure could potentially trigger hundreds of alerts across dozens of tables and multiple domains or teams.

This approach can spark inefficient, overlapping fire drills without clear prioritization of the most urgent issues. Worse yet, it can lead teams to experience alert fatigue and, eventually, apathy.

Data observability: integrated lineage & intelligent grouping

Data observability platforms leverage integrated data lineage to provide context for alerts. With a clear map of how upstream sources and downstream assets are related, the platform can group related issues intelligently, conduct automatic impact analysis, and provide important context about severity within its intelligently routed alerts. The result? Quicker insights into the impact and point of origin of data issues. This makes ownership clear and incident response more efficient, reducing alert fatigue and creating a culture of urgency around truly impactful problems.



Resolution: Automated Fixes vs. Root Cause Analysis

How each type of solution approaches resolution is probably the most significant difference between traditional data quality solutions and data observability platforms — pay close attention here.

Traditional data quality solutions: automated fixes with hidden risks

As a general rule, data quality solutions do not provide any support for root cause analysis.

However, some data quality tools do offer features that directly remediate or transform data, such as automatically deleting duplicate entries or quarantining suspect data.

But proceed with caution — automated fixes can potentially create new errors or mask underlying issues without addressing their root cause. This approach treats symptoms rather than solving core problems, allowing faulty data pipelines to continue operating unchecked.

Data observability: efficient yet thorough root cause analysis

Data observability platforms take a different tack, focusing on empowering human decision-making rather than automated fixes. These solutions provide robust support for root cause analysis, correlating anomalies to their sources across data, systems, and code.

This approach allows teams to resolve data incidents efficiently at scale by providing visibility into the three primary root causes of data anomalies: problematic data sources, ELT system or pipeline failures, and mistakes in transformation code.
And by keeping humans in the loop, it ensures that real problems are fixed at their source — preventing future issues and building institutional knowledge.

Measurement: Rule Breaches vs. Holistic Reliability

Both traditional data quality solutions and data observability platforms consistently measure data health — but across wildly different scopes.

Traditional data quality solutions: focused on rule violations

Data quality solutions usually feature scorecards that surface the number of breached rules in a table. This provides a snapshot of the current data state, but fails to answer questions like: How long has this rule been breached? How quickly are incidents resolved? What are the reliability and coverage levels upstream?

In other words, data quality measurement doesn't actually communicate whether a data product can be trusted.

Data observability: ensuring holistic reliability

Data observability takes measurement to the next level. In addition to simple scorecards like the ones data quality tools provide, data observability dashboards report on reliability at the organizational, domain, data product, and table levels. This includes tracking how long rules have been breached, incident resolution times, and upstream reliability and coverage levels.

The result is a comprehensive view of data health that goes beyond simple rule violations. By providing insights into the uptime of data products and the operational response of teams, data observability offers a more nuanced and actionable perspective on data reliability.

Cost Optimization: Limited Insights vs. Comprehensive Analysis

Managing data costs is a huge concern for modern data teams, as storage and compute bills can spiral out of control without careful monitoring and management. But data quality solutions and data observability platforms offer very different capabilities when it comes to cost optimization.

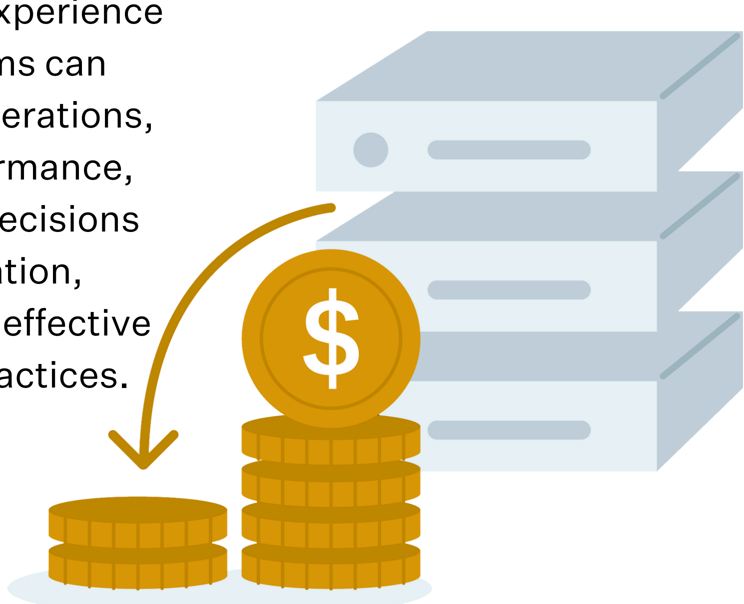
Traditional data quality solutions: limited cost visibility

Again, data quality solutions are focused on the content of the data itself. They typically don't provide insights into job runtime or cost optimization, leaving teams in the dark about the financial implications of their data operations. This means teams may unknowingly run inefficient, expensive operations, leading to inflated data management costs.

Data observability: actionable cost insights

Data observability platforms, on the other hand, are breaking new ground in cost optimization. By providing insights into the runtime of jobs executed on modern data platforms, these tools allow teams to understand and optimize their data operations from a financial perspective.

The result is the ability to balance cost, reliability, and user experience simultaneously. Teams can identify expensive operations, optimize query performance, and make informed decisions about resource allocation, leading to more cost-effective data management practices.



Are Data Quality Solutions & Data Observability Platforms Better Together?

As we’ve covered at length, data quality solutions and data observability platforms go about their common goal — improving data quality — in very different ways. But since this is such an important area of focus for data teams, isn’t it better to cover all your bases?

Not exactly. It’s not bad to have both kinds of data management tools in place — but in some cases, it’s also necessary. In most cases, a comprehensive data observability platform will meet an organization's data quality and reliability needs.

What Analysts Have to Say

Most industry analysts agree that augmented data quality solutions alone will not get the job done.

- "Data quality is the result of powerful data observability across the modern data stack." - Shalaka Joshi, G2 Market Research Analyst
- "Enterprises that embrace data observability have the potential to improve the quality of the data as it is generated and processed, as opposed to checking for quality problems after the event." - Matt Aslett, Ventana Research Analyst
- "Static, event-based monitoring is no longer sufficient to effectively manage data systems and prevent critical events. Modern data architectures are too complex and dynamic for those methods to provide a holistic view of data's health across data ecosystems at various stages of its life cycle." - Gartner

Again, we can look at the parallels between data observability and application observability to understand the nuance here. In APM, software engineers supplement unit tests with observability solutions to handle increasing application complexity in the cloud. Similarly, data organizations must supplement data quality monitoring with data observability to ensure the reliability of data products in complex modern data architectures.

If you already are using data quality solutions, especially if different teams already have their own tools in place, they should all be integrated with a data observability platform to avoid silos and provide a more holistic view of data health.

Ultimately, the key is to bring engineers, analysts, and governance managers together to look at the full context of data reliability. Building a culture of data trust requires collaboration and visibility, and a data observability platform can help facilitate a new level of data-driven decision-making.

So let's take a look at how some real-life data teams have done exactly that.

Real Stories of Data Quality in the Enterprise

JetBlue: Elevating Data Reliability with Data Observability

Prior to adopting data observability, commercial airline JetBlue grappled with data trust issues following their migration to Snowflake. Their modern data stack, encompassing 3,400 analyst-facing tables and views across 5 petabytes of data, faced increased scrutiny as more data sets were integrated into their ecosystem. Traditional data quality measures were falling short in this complex environment.

By implementing Monte Carlo's data observability platform, JetBlue transformed their approach to data reliability:

- AI-First Detection: The platform deployed AI-driven monitoring for volume, freshness, and schema changes across all Snowflake tables, moving beyond rule-based detection to catch unforeseen issues.
- Integrated Triage: JetBlue leveraged intelligent alert grouping and automatic impact analysis, enabling quicker identification of issue origins and more efficient incident response.
- Root Cause Analysis: The data operations team evolved from merely monitoring pipeline failures to proactively addressing data anomalies, tracing issues across data, systems, and code.
- Comprehensive Measurement: JetBlue implemented six key metrics to assess data observability success, including incident classification rate and time to resolution, providing a holistic view of data health.
- Data Trust as a KPI: The airline saw their Data NPS score increase by 16 points year-over-year, quantifiably demonstrating improved data trust across the organization.

This shift from traditional data quality management to comprehensive data observability allowed JetBlue to not only enhance data reliability but also foster a culture of proactive data stewardship.

“When Monte Carlo identifies an incident, it shows you enough metadata so that you’re able to understand what’s being impacted down the line, including how many people are potentially being impacted by the issuer.” - Brian Pederson, Manager of Data Products, JetBlue

Read JetBlue’s full story of adopting data observability.

jetBlue®

Roche: Transforming Data Management with Data Mesh and Observability

Before embracing data observability, biotech firm Roche operated on legacy on-premises infrastructure with a classic monolithic architecture. This setup resulted in slow release cycles, averaging three months, and difficulties in scaling compute resources. The data team scored low on reliability and performance satisfaction indicators, with multiple major data quality incidents occurring every year.

Roche's adoption of data observability, as part of a broader data mesh strategy, brought about significant transformations:

- Holistic data visibility: The data team implemented a comprehensive data observability approach with Monte Carlo, covering data content, pipelines, and infrastructure.
- Data as a product: By treating data as a product, Roche’s data team shifted from reactive data quality checks to proactive data reliability management, emphasizing the importance of end-to-end visibility.
- Automated measurement: Roche implemented automated dashboards to track data product health, including freshness, volume, schema, and quality metrics, providing a comprehensive view of data reliability for self-serve users.

By implementing this data observability framework within their data mesh architecture, Roche transformed their data management approach. They moved from a slow, reactive model to a proactive, self-service oriented system that prioritizes data reliability and trust.

Follow Roche’s complete journey to a data mesh.

Fortune 500 Global Telecommunications Company: Scaling Data Reliability Across Billions of Daily Transactions

A Fortune 500 global telecommunications company's digital advertising division faced the challenge of managing data quality at an enormous scale — processing about 5 billion transactions each month where viewers engage with advertisements, enriched with data from 20 additional partnership sources.

Before implementing data observability, the company struggled with the reliability of data powering critical dashboards that tracked advertising campaign performance. Data quality issues could potentially impact thousands of customers and millions of dollars in revenue. And traditional data quality measures were insufficient for their complex, high-volume data environment.

After adopting a comprehensive data observability approach, they realized:

- End-to-End Visibility: By integrating data observability with their data catalog, the company gained full visibility into data flows across their tables, enabling efficient tracing of issues to their source.
- Automated Monitoring: The data observability platform quickly scaled monitoring for data freshness, volume, and schema across their environment and production pipelines, delivering value within weeks.
- Custom Monitoring: The team implemented field health monitors on critical tables and created custom SQL-based monitors for key internal metrics, ensuring comprehensive coverage of their specific needs.
- Intelligent Alerting: A smart notification strategy was designed, categorizing incidents by severity and domain, and routing them to appropriate teams via Microsoft Teams.
- Root Cause Analysis: Data lineage features played a central role in the remediation process, helping teams understand data flows and trace issues to their source.

These powerful capabilities help the digital ad team ensure they’re displaying the latest campaign pacing metrics to customers, monitor real-time campaign performance at a granular level, and troubleshoot effectively when issues occur. The big-picture results include improved efficiencies across teams, higher confidence in data-driven decision making, and enhanced reliability of critical reports.

Embracing the Future of Data Quality

Data quality solutions have been an integral part of the data stack for decades — but as the landscape of data management keeps growing in complexity and scale, these traditional approaches simply aren’t enough to safeguard data health across its full lifecycle. Data observability expands the scope of data quality to the entire lifecycle, providing visibility and facilitating proactive improvements to data health across the entire data stack.

Leading data teams at organizations like JetBlue, CreditKarma, SeatGeek, HubSpot, Roche, Fox, PepsiCo, and more are reaping the benefits of data observability: improving data quality, reducing incident response times, and building a culture of data trust.

Don't wait for the next data crisis to strike. Proactively safeguard your data assets and empower your team with automated, end-to-end data observability. Contact our team and take the first step towards true data reliability.

Additional Resources

Check out more helpful resources on data and AI trends and best practices, including:

- Data Downtime Blog: Get fresh tips, how-tos, and expert advice on all things data.
- O'Reilly Data Quality Framework: The first several chapters of this practitioner's guide to building more trustworthy pipelines are free to access.
- Data Observability Product Tour: Check out this video tour showing just how a data observability platform works.
- Data Quality Value Calculator: Enter in a few specifics about your data environment and see how much you can save with data observability.