MONTE CARLO

The State of Data Quality for 2026

Data based insights into operational best practices from monitoring 11m tables.

TABLE OF CONTENTS

Introduction

The Data Quality Paradox

Every day, more than 1,000 data quality incidents are detected and resolved in the Monte Carlo Data + AI Observability platform.

Across hundreds of organizations and millions of tables, thousands of data engineers, analysts, and AI practitioners rely on Monte Carlo to monitor, detect, and resolve reliability issues before they impact the business.

This scale gives us a unique vantage point into how modern teams approach reliability, the most common causes of bad data, and what separates top-performing teams from the rest. Adopting data quality best practices isn't just about cleaner pipelines — it's about trust. Reliable data drives accurate dashboards, dependable AI models, and confident decisions. Unreliable data erodes that trust just as quickly.

This report aims to uncover:

- How data quality varies across modern tech stacks
- The most common issues degrading reliability
- How top-performing data teams are closing the trust gap

The Data Reliability Problem

How Often Does Bad Data Strike?

To understand the state of data quality, we looked deeply into telemetry across our customer environments. Our goal: estimate how many significant data reliability issues the average team encounters in a given year.

The results were striking — and sobering.

- When we first measured this in 2020, we found roughly 1 data quality issue for every 15 tables per year.
- By 2023, despite improved tooling and awareness, that number has increased to 1 issue for every 10 tables.

> **1 in every 10 tables experiences a data quality issue per year.**

This means if your organization has 10,000 tables in production, you can expect around 1,000 incidents each year — any of which could delay reports, disrupt machine learning pipelines, or mislead critical decisions.

Estimating Data Downtime

We define data downtime as the period when data is missing, inaccurate, or otherwise unusable. Using our benchmark, you can estimate your organization's annual data downtime with this formula:

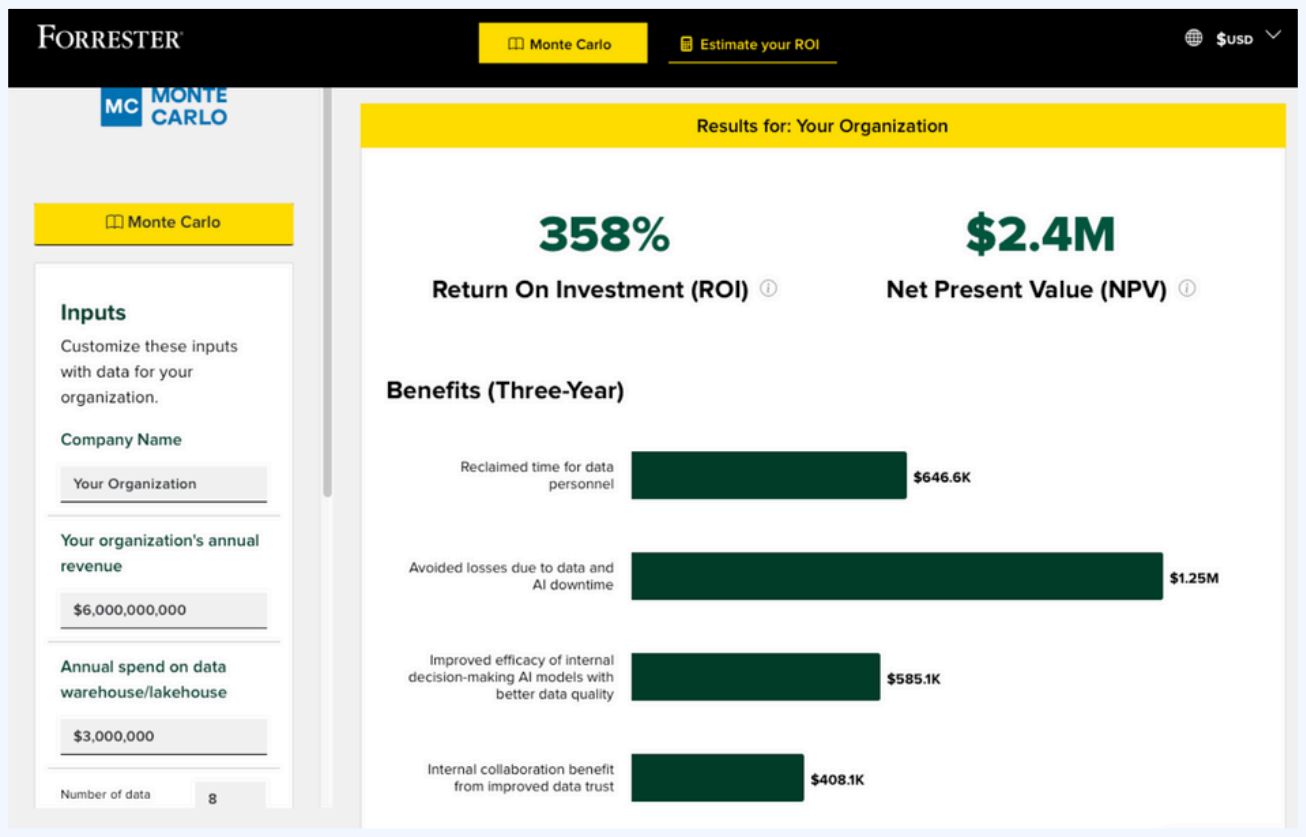
> **(Total Tables ÷ 10) × 15 hours = Estimated annual data downtime**

That 15-hour figure comes from a 2023 survey of 200 data professionals, who reported that it takes an average of 15 hours to detect, investigate, and resolve each incident.

For a data environment of 10,000 tables, that's 15,000 hours — nearly two years of lost productivity every year.

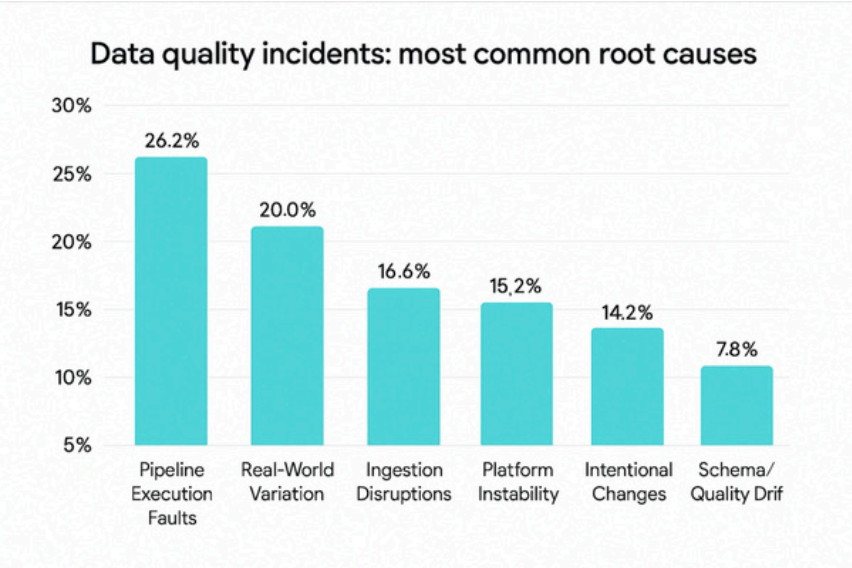
To put that in perspective, that's thousands of hours of delayed insights, misinformed decisions, and frustrated business users.

To explore the cost of this downtime, Monte Carlo partnered with Forrester to develop a Data Downtime Calculator. This tool helps teams estimate how much unreliable data costs in lost revenue, efficiency, and trust.



What Causes Poor Data Quality?

Monte Carlo's Troubleshooting Agent has been deployed thousands of times across hundreds of customer environments to pinpoint the root causes of data quality incidents. When we aggregate this telemetry, a clear story emerges: data quality issues aren't the result of a single failure — they're the outcome of a complex ecosystem of interdependent systems, human processes, and fragile pipelines.



Data quality issues aren't the result of a single failure — they're the outcome of a complex ecosystem.

The Takeaway

More than one-third (34%) of incidents aren't "true" errors — they're expected variations or intentional changes. The challenge lies in identifying which incidents are real problems worth acting on.

> **Observability is the ability to have visibility into the inputs and outputs of a system as well as the performance of its component parts.**

The brittleness of enterprise data systems — from orchestration tools to vendor APIs — means reliability requires more than validation rules. It demands observability, visibility, and intelligent context.

Incident Management (Triage & Resolution)

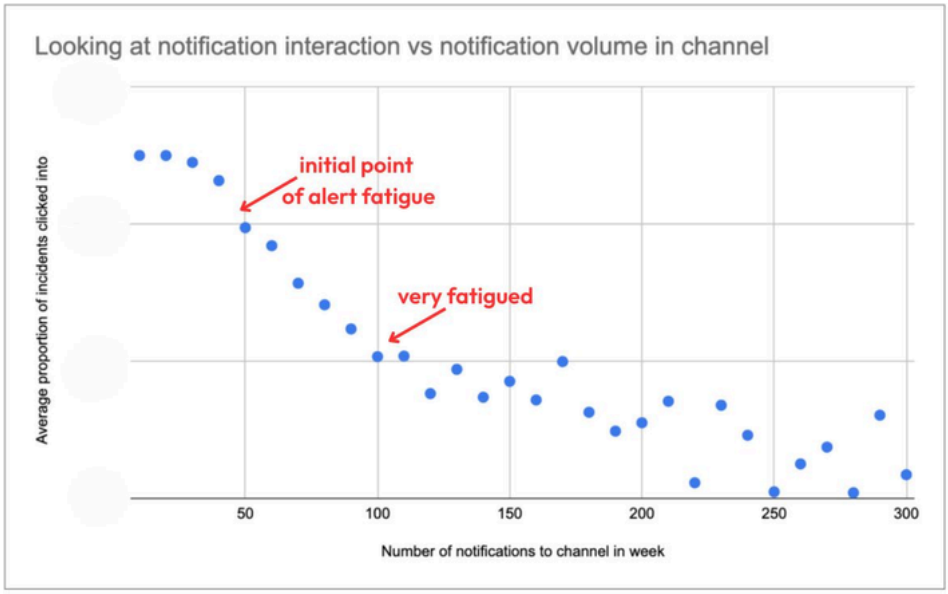
Once a data quality issue is detected, how a team responds determines the impact. Through thousands of incident logs, we've identified four key principles that drive effective triage and resolution:

1. Don't overwhelm teams with too many alerts in one channel.
2. Route alerts to the right people through the right systems.
3. Designate clear ownership for incidents.
4. Involve leadership in reviewing data reliability trends.

Let's break these down.

Alert Routing

Alert fatigue is one of the biggest barriers to reliability. Our research shows that when **a channel receives more than 50 alerts per week**, the rate of updates and **responses drops by 15%**. When **alerts exceed 100 per week**, engagement **falls another 20%.**



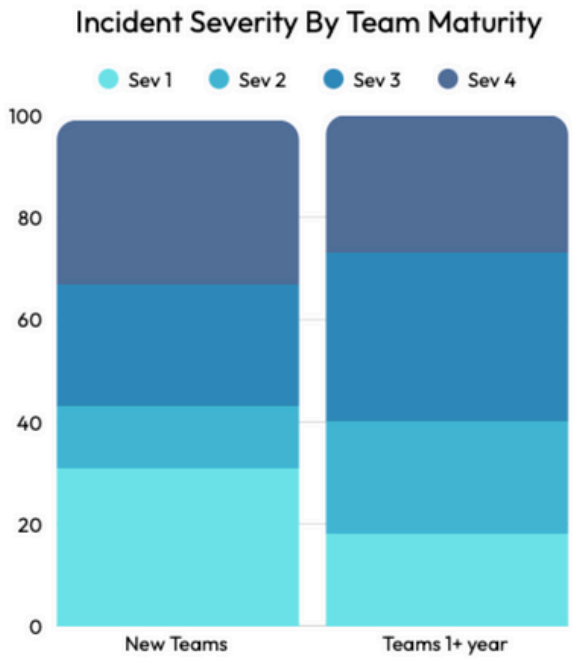
Teams that route alerts to Slack or persistent chat channels see a 4x higher click-through rate than email. Even more interestingly, alerts sent to incident management tools like JIRA or OpsGenie perform 10% better than Slack, likely because they reach on-call, high-priority responders directly.

Incident Severity

When every issue feels urgent, nothing is urgent. In the early stages of implementing data reliability, we see teams label 32% of all incidents as Sev 1 (critical).

As maturity improves — and teams define consistent severity rules — Sev 1 incidents drop to 18%, while lower-severity classifications increase proportionally.

This normalization reflects improved process control and more realistic prioritization — hallmarks of a mature data organization.



Incident Collaboration

Data reliability is a team sport, but not all players have equal impact.

In Monte Carlo's telemetry, 20% of users escalate 80% of alerts, reinforcing the classic 80/20 rule. Most alerts have only one active participant, and surprisingly, incidents with multiple participants actually take longer to resolve — because complex issues often require collaboration across multiple domains. Interestingly, incidents with designated owners are resolved 1.5x faster on average, but take longer to close — likely because owners handle the most critical, complex cases.

Executive Involvement

The strongest data reliability cultures are championed from the top. Teams whose executives actively review data health dashboards achieve 1.5x higher engagement rates and faster response times. Organizations like JetBlue have institutionalized this, reviewing every alert at the leadership level on a biweekly cadence.

> **"When leaders engage with data quality, teams respond with accountability."**

Monitoring and Automation

Effective monitoring is the foundation of data reliability.

Analyzing the performance of thousands of monitors, we found that automation and placement make a measurable difference. Teams using anomaly detection monitors report 40% fewer manual updates than those relying on SQL or static validation.

The tradeoff: fewer manual interventions, but slightly lower engagement due to reduced human tuning.

Monitor Type	Avg. Touches	Feedback Rate
Custom SQL	2.35	High (30% above baseline)
Data Validation	2.03	Moderate
Comparison	1.63	Moderate
Anomaly Detection	1.33	Lower (but more efficient)

Monitor Placement

The most effective teams monitor end-to-end across pipelines. Monte Carlo data shows:

- 34% of monitors are at the landing (bronze) layer
- 50% at the transform (silver) layer
- 26% at the gold / serving layer

This distribution makes root-cause analysis dramatically faster — when every layer of the data lifecycle is visible, engineers can pinpoint where an incident begins.

Building a Culture of Reliability

Troubleshooting isn't just about fixing what's broken — it's about designing for resilience.

Teams that excel at reliability share a few defining traits:

- They treat data observability as an operational discipline, not a reactive fix.
- They establish clear ownership for incidents and maintain accountability loops.
- They automate wherever possible — 80% of disruptions can be reduced with intelligent observability and root cause correlation.
- They invest in cross-functional collaboration between data engineers, analysts, and leadership.

> **"Reliable data isn't built by chance — it's built by culture."**

Going Beyond Data Quality

Modern data teams are realizing that achieving trust requires more than clean tables or error-free dashboards. It demands a system-wide commitment to observability, accountability, and continuous improvement.

The takeaway is clear: Data downtime isn't inevitable — it's manageable. By monitoring end-to-end pipelines, improving alert hygiene, assigning clear ownership, and engaging leadership, organizations can dramatically reduce both the frequency and cost of bad data.

But the most forward-thinking teams go further. They build reliable data ecosystems — systems that not only detect and fix issues, but also learn, adapt, and prevent them from happening again.

Monte Carlo's Data + AI Observability platform empowers teams to go beyond data quality — helping them detect, resolve, and prevent reliability issues at scale, so every decision, model, and product is built on data the business can trust.

Accelerate Resolution by 80%+ with Troubleshooting Agent

I hope you've found this guide useful. If I did accomplish my goal to provide insightful and trusted content based on real-life experiences, please feel free to give me a shout on LinkedIn and tell me what you found most helpful or what other questions can be covered.

If you are interested in learning more about Monte Carlo's data + AI observability solution, we'd love to talk with you.

Request a demo