

The 2026 Enterprise Playbook for Data Observability:

A Strategic Framework for AI-Ready Data



Trust Your Data.

Control Your Costs.

Power Your AI.

Table of Contents

Publication Information

Published by: DataRadar, LLC

Written by: Ken Kasee

Publication Date: January 2026

Contact Information:

Website: www.dataradar.io

Email: info@dataradar.io

LinkedIn: [linkedin.com/company/dataradar-io](https://www.linkedin.com/company/dataradar-io)

Legal Notice

This playbook is intended for educational purposes and embodies industry best practices and research as of January 2026. It does not constitute legal, compliance, financial, or other professional advice. Users should consult qualified professionals before making decisions based on the Playbook's contents. While it references DataRadar capabilities, its frameworks, assessments, and recommendations are valuable irrespective of technological choices.

All statistics and research are cited with footnotes linking to sources. This contains proprietary research and analysis. No part of the Playbook may be copied, reproduced, modified, adapted, translated, publicly displayed, performed, distributed, transmitted, sold, licensed, or otherwise exploited in any form or by any means, including without limitation electronic, mechanical, photocopying, recording, or other methods, without the prior written permission of DataRadar, Inc. Reproduction without written permission is prohibited. © 2026 DataRadar, LLC. All rights reserved.

You may not use any trademarks, service marks, logos, or trade dress appearing in the Playbook without the express prior written consent of their respective owners. All rights not expressly granted are reserved by DataRadar, Inc.

DataRadar, LLC, may pursue all remedies available at law or in equity for any violation of this Notice or infringement/misappropriation of its intellectual property rights.

©2026 Data Radar, LLC.
All rights reserved.

How to Use This Playbook	3
What Is This Playbook?	3
Who Should Use This Playbook?	3
How to Navigate the Playbook	4
Executive Summary	5
What You'll Learn	5
Chapter 1: The \$12.9M Problem	6
Why AI Initiatives Fail	7
The Economics of Bad Data	8
The Predictable Failure Pattern	9
Six Critical Data Quality Dimensions for AI	10
The Observability Imperative	10
Executive Summary: Chapter 1	11
Questions to Ask Your Organization	11
Chapter 2: Nine Trends Reshaping Data Operations	12
Introduction	13
Trend 1: AI-Powered Predictive Observability	14
Trend 2: Cost-Aware Data Tiering and FinOps Integration	15
Trend 3: Unified Data and AI Observability	19
Trend 4: Observability for Agentic AI Governance	21
Trend 5: Observability as Business Value Driver	23
Trend 6: Data Productization and Self-Service Analytics	26
Trend 7: Real-Time Data Quality Monitoring	28
Trend 8: Cloud-Native Observability Architectures	30
Trend 9: Regulatory Pressure and Compliance Automation	33
The Regulatory Landscape	33
Executive Summary: Chapter 2	35
Questions to Ask Your Organization	36
Chapter 3: The 90-Day Implementation Roadmap	37
Overview	38
The AI-Ready Data Maturity Model	38
Phase 1: Assessment & Foundation (Days 1-30)	40
Phase 2: Quick Wins & Visibility (Days 31-60)	40
Why Native Architecture Matters	41
Phase 3: Scale & Optimize (Days 61-90)	42
Success Metrics	42
Executive Summary: Chapter 3	43
Questions to Ask Your Organization	43
Chapter 4: The Agentic AI Era—Future-Proofing for 2027 and Beyond	44
The Agentic AI Revolution	45
Autonomous Data Operations	46
The Multi-Modal Data Challenge	47
Principles That Endure	48
Executive Summary: Chapter 4	49
Questions to Ask Your Organization	50
Conclusion: Building the Foundation for AI Success	53
The Path Forward	53
Appendix A: Complete Source Citations	54
Appendix B: Glossary of Terms	59
Appendix C: About DataRadar	61
The DataRadar Difference	62
Contact Information	62

How to Use This Playbook

What Is This Playbook?

This document serves not merely as a guide but as an actionable strategic manual for data leaders navigating the most significant transformation in data operations since the transition to cloud technology. Consider it your comprehensive roadmap, integrating:



Industry Research:

Insight from the leading analysts on data observability



Practical Frameworks:

The AI-Ready Data Maturity Model and a 90-day implementation roadmap



Future Perspective:

Nine interconnected trends shaping 2026 and beyond



Self-Assessment Tools:

Questions to evaluate your organization's readiness



Who Should Use This Playbook?

Primary Audience:

- Chief Data Officers and VP-level data leaders making strategic decisions
- Data Engineering Directors implementing technical solutions
- AI/ML Program Leaders ensuring data readiness for AI initiatives

Secondary Audience:

- FinOps Teams connecting data quality to cloud costs
- Data Governance Officers managing compliance and risk
- IT Executives evaluating platform architectures

How to Navigate the Playbook

For Executives (30-minute read):

1. Read the Executive Summary and **Chapter 1** (The \$12.9M Problem)
2. Skim **Chapter 2** headlines to understand the nine trends
3. Review the 90-Day Roadmap in **Chapter 3**
4. Focus on **Executive Summaries** and 'Questions to Ask' sections



For Practitioners (2-hour deep dive):

1. Complete reading of **all chapters**
2. Use **self-assessment questions** to evaluate the current state
3. **Map the 90-day roadmap** to your organization's priorities
4. Reference specific trends as you **plan initiatives**

For Strategic Planning:

1. Use **Chapter 1** to build the business case
2. Apply the Maturity Model in **Chapter 3** for gap analysis
3. Leverage **Chapter 2** trends for roadmap planning
4. Reference **Chapter 4** for future-proofing decisions



Executive Summary

The data landscape has undergone a fundamental transformation. Organizations face a significant challenge: 88% of AI pilot projects do not progress to production, with insufficient AI-ready data cited as a leading obstacle.¹ The financial impact of poor data quality averages \$12.9 million per year, with larger enterprises losing \$15 million or more annually.²

This playbook offers enterprise data leaders a strategic framework for building reliable, AI-ready data foundations through modern observability practices. By 2028, the most successful organizations will not be defined by the most advanced AI models; rather, they will be distinguished by the most trusted data.

Executive Summary Key Stats Dashboard

	\$12.9M	Average annual cost of poor data quality per organization.
	88%	Of AI projects fail to reach production due to data issues.
	70%	Of enterprises will adopt data observability tools by 2027.

Data quality is expensive, AI projects are failing, but there's a clear path forward.

- Trust Your Data.**
- Control Your Costs.**
- Power Your AI.**

This principle influences all trends, strategies, and recommendations outlined in this playbook. Contemporary data operations require a unified approach that enables simultaneous consideration of quality and cost, observability and optimization, and governance and agility.



What You'll Learn

- Why AI initiatives fail and the hidden costs of poor data quality
- Nine interconnected trends reshaping enterprise data operations in 2026
- How unified platforms solve fragmented tool sprawl
- A practical 90-day implementation roadmap
- Future-proofing strategies for emerging technologies

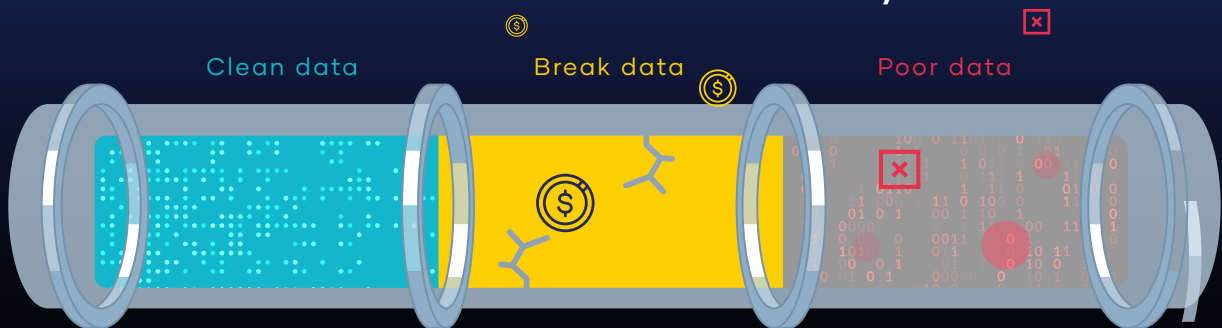
The background features a dark teal-to-blue gradient with a faint grid pattern. Overlaid on this are numerous glowing, semi-transparent geometric shapes, including cubes, rectangles, and lines, some of which are interconnected to form a complex, abstract structure. The overall aesthetic is technical and digital.

Chapter 1: The \$12.9M Problem

Why AI Initiatives Fail

In boardrooms around the world, a familiar pattern unfolds. Leadership teams enthusiastically announce exciting AI initiatives such as predictive maintenance, smarter underwriting, and more personalized customer experiences. Data science experts work diligently to develop advanced models. The initial results are often encouraging, sparking optimism. However, sooner or later, the challenges and complexities become clear, reminding us that implementation can be more intricate than it first appears. Models tend to degrade over time, making predictions less reliable. This can erode business users' confidence. Sadly, such initiatives can often become just another frustrating addition to the growing list of unsuccessful AI projects.

CHAPTER 1 OPENER: The Cost of Poor Data Quality



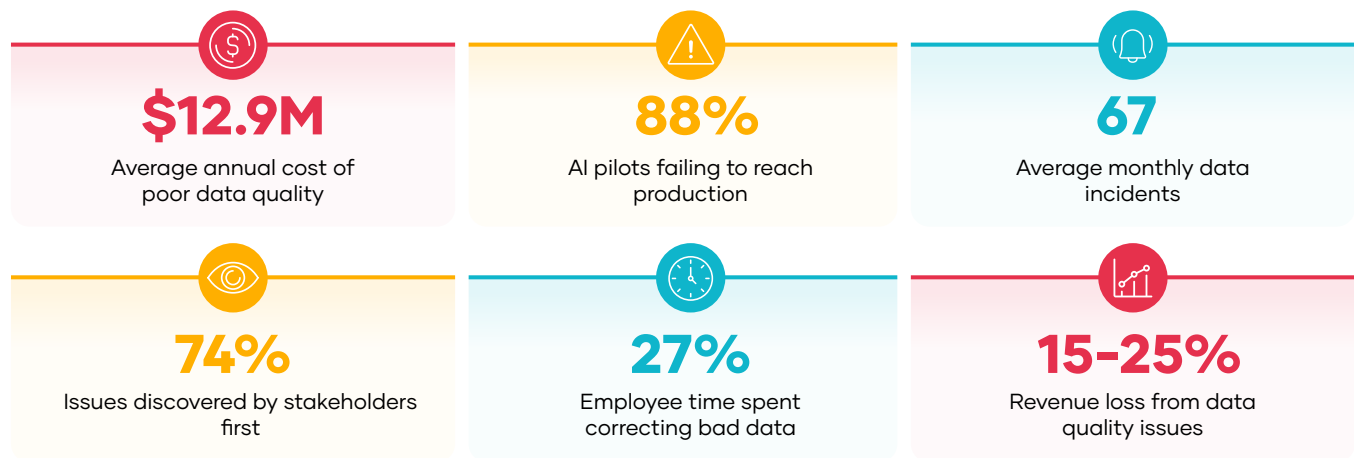
Poor data quality silently drains value from AI investments. The loss happens gradually at multiple points.

**This isn't just about a technology hiccup;
it's about a data issue.**

The Economics of Bad Data

The financial impact of poor data quality is staggering. Research shows that only 12% of organizations report that their data is of sufficient quality for effective AI implementation, and 50% cite data quality as the number one issue affecting their data integration projects.³

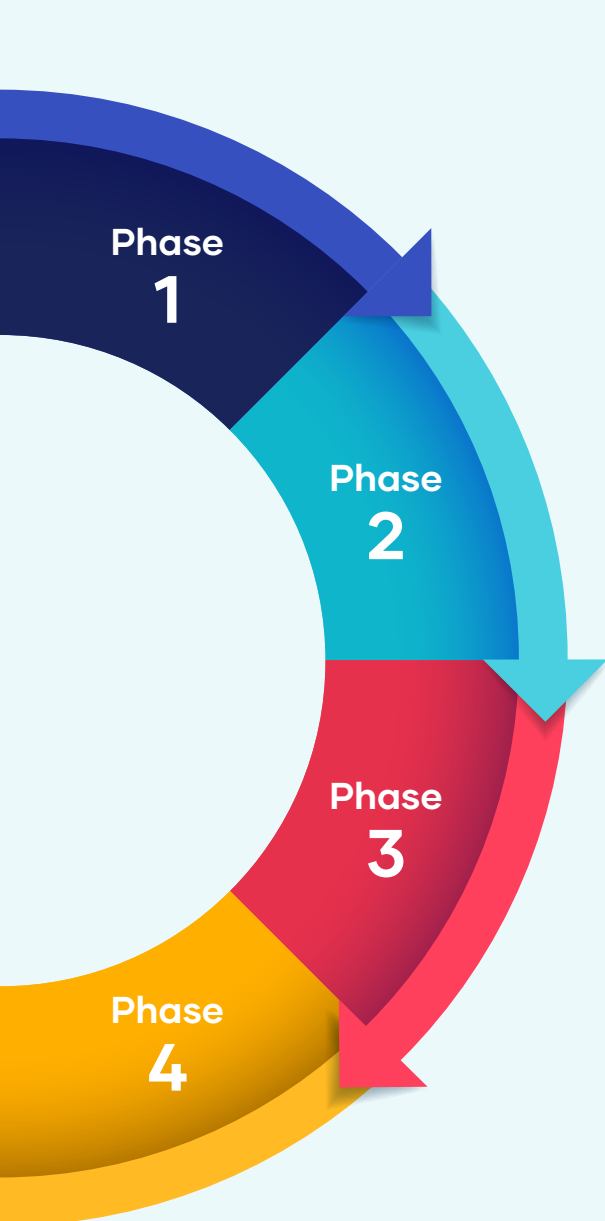
Key Statistics Dashboard



Metric	Value
Average annual cost of poor data quality	\$12.9M per organization Gartner, 2021 ⁴
AI pilot projects failing to reach production	88% DC, 2024 ⁵
Organizations that don't measure data quality costs	59% Gartner ⁴
Data quality sufficient for effective AI implementation	Only 12% Precisely/Drexel University, 2024 ²
Analyst time spent vetting and validating data	>40% (for nearly 1/3 of analysts) Forrester ⁶
Revenue loss attributed to data quality issues	15-25% MIT Sloan Management Review ²

The Predictable Failure Pattern

AI initiatives face a sobering reality: for every 33 proof-of-concept launches, only 4 reach production.¹⁰ RAND Corporation research confirms that over 80% of AI projects fail overall—twice the rate of traditional IT projects.¹² This pattern typically unfolds in four phases:



Phase 1 Enthusiasm

Teams build a proof-of-concept on a curated dataset. The model performs excellently in controlled conditions, and optimism runs high. Executives greenlight expansion.

Phase 2 Reality

The model encounters production data and immediately faces missing values, schema changes, duplicate records, and data drift. Performance degrades rapidly. The gap between prototype and production becomes painfully clear.

Phase 3 Firefighting

Data engineers rush to fix issues reactively. The team becomes trapped in constant triage, unable to improve the model because they are too busy keeping it alive. Leading analysts found that 85% of AI projects fail due to poor data quality at this stage.¹³

Phase 4 Abandonment

Business stakeholders lose patience. The AI initiative is deprioritized, shelved, or quietly canceled. S&P Global reports that 42% of companies abandoned most of their AI initiatives in 2025, up from just 17% in 2024.¹⁴

This pattern is not inevitable. Organizations with proper data observability break the cycle by detecting issues before they cascade, shifting from reactive firefighting to proactive prevention.

Six Critical Data Quality Dimensions for AI

The Data Management Association (DAMA) defines six core dimensions of data quality: accuracy, completeness, consistency, timeliness, validity, and uniqueness.¹⁵ For AI and machine learning initiatives, leading research extends this framework with two additional observability dimensions essential to production ML systems:

1. **Freshness** measures how recently data was updated. Stale data leads to outdated decisions. A fraud detection model using week-old transaction data will miss emerging patterns.
2. **Completeness** assesses whether all expected values are present. Missing values force models to impute data, introducing errors that compound over time.
3. **Consistency** evaluates whether values align across sources. When the same customer appears with different addresses in different systems, models become confused and predictions suffer.
4. **Accuracy** determines whether values reflect reality. Wrong data produces wrong predictions and wrong business decisions.
5. **Schema Stability** tracks whether data structure remains predictable. When upstream systems change schemas without warning, pipelines fail and models break. This observability-specific dimension is critical for production ML.
6. **Lineage** enables tracing data back to its source. Without lineage, debugging becomes impossible. When a model produces unexpected results, teams cannot identify which data caused the problem.

Organizations that monitor all six dimensions catch issues before they affect production models.

The Observability Imperative

Traditional approaches to data quality are reactive. Teams discover problems only after damage occurs, often when a stakeholder complains. Modern data observability goes beyond traditional monitoring and detection, alerting teams to issues before they escalate.¹⁶

This shift manifests in three key areas:

1. From sampling to full coverage. Traditional approaches rely on statistical samples that may miss edge cases. Modern observability monitors 100% of data, ensuring no anomaly goes undetected.
2. From batch processing to real-time detection. Traditional monitoring runs in nightly jobs. Modern observability detects issues as they happen.
3. From manual rules to intelligent detection. Traditional monitoring depends on human-defined rules. Modern observability uses algorithms that learn normal patterns automatically.

Organizations with full-stack observability experience 79% less downtime and are 51% more likely to detect interruptions earlier than those without it.¹⁷ The true benefit comes from prevention, identifying and addressing issues before they propagate throughout your data ecosystem.

Executive Summary: Chapter 1



The Core Problem:

88% of AI pilots fail to reach production, primarily due to data quality issues costing organizations an average of \$12.9M⁴ annually.



Critical Dimensions:

Six data quality dimensions determine AI success: Freshness, Completeness, Consistency, Accuracy, Schema Stability, and Lineage.



Key Insight:

AI does not fail because of bad algorithms—it fails because of bad data. The predictable failure pattern moves from enthusiasm through reality to firefighting to abandonment.



The Solution:

Modern data observability provides proactive, intelligent detection and prevention rather than reactive cleanup.



Questions to Ask Your Organization

Strategic

1. What percentage of our AI initiatives have successfully moved from pilot to production?
2. Do we know the actual cost of data quality issues to our organization?
3. Who owns the data quality agenda, and do they have executive sponsorship?

Tactical

1. How are data quality problems typically discovered by our team proactively, or by stakeholders complaining?
2. What is our current mean time to detect (MTTD) and mean time to resolve (MTTR) for data incidents?
3. What percentage of our time is spent on reactive firefighting versus proactive improvement?

Financial

1. Are we treating data infrastructure as a cost center or a strategic enabler?
2. Do we have documented data quality standards and SLAs that are actively tracked?



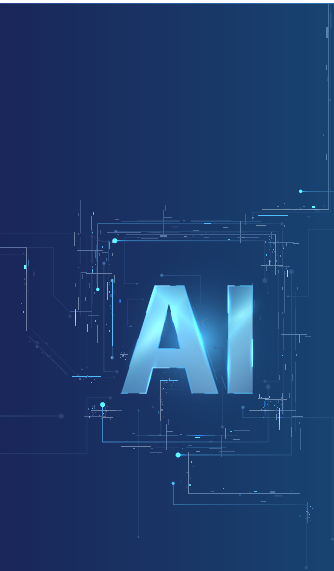
Chapter 2: Nine Trends Reshaping Data Operations

Introduction

Something fundamental is shifting in how organizations think about data infrastructure.

For years, data observability was a niche concern. Data engineering teams monitored pipelines, fixed issues, and moved on. The work was necessary but invisible. Leadership viewed it as plumbing: important when it failed, ignored when it worked.

That era is ending. The rise of AI, the pressure to control cloud costs, and the acceleration of regulatory requirements have elevated data observability from an operational necessity to a strategic priority. What happens in your data pipelines now directly affects whether AI initiatives succeed, whether cloud budgets stay under control, and whether your organization remains compliant with an expanding web of regulations.



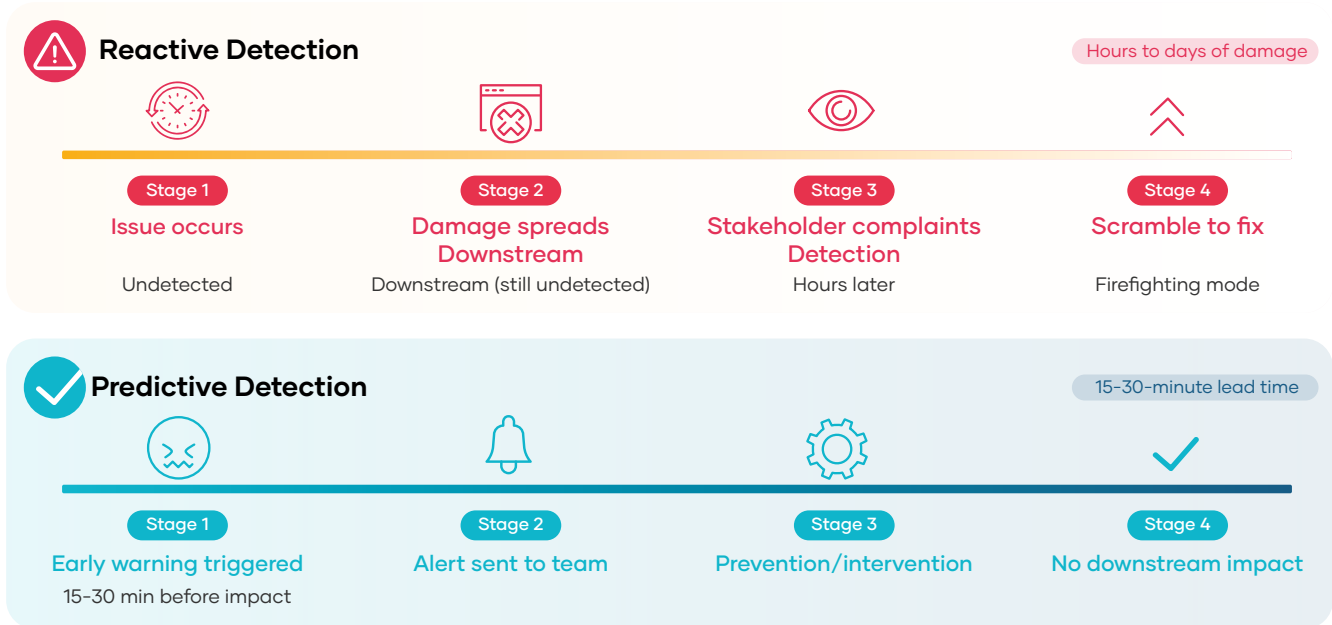
The numbers reflect this shift. The global data observability market reached \$3.2 billion in 2025, with forecasts projecting growth to \$5.45 billion by 2030¹⁸. Leading analysts project that 50% of enterprises implementing distributed data architectures will adopt data observability tools by 2026, up from less than 20% in 2024.¹²

But market growth alone does not capture what is happening. The nature of observability itself is changing. Tools that once simply detected problems are now predicting them. Platforms that once focused solely on data quality are now optimizing costs. Systems that once served only technical teams are now delivering business metrics to executive dashboards.

These nine trends are not isolated phenomena. They are interconnected components of a fundamental transformation in how organizations generate, oversee, and derive value from data. Understanding them is essential for any organization building an AI-ready data foundation.

Trend 1: AI-Powered Predictive Observability

Predictive vs. Reactive Timeline



15-30 min advantage Predictive detection gives teams 15-30 minutes of lead time to prevent damage that reactive approaches only discover after the fact.

From Reactive to Predictive

Traditional monitoring lets you know about issues only after they occur. Predictive observability, by contrast, uses machine learning to anticipate problems in advance, helping prevent disruptions to your systems and processes.

Analyst research shows that companies with effective observability can reduce system downtime by up to 50% and improve incident response time by 40%.²⁰ But speed is not everything. What truly matters is preventing issues before they occur, adding real value to their efforts.

How It Works

Predictive observability systems keep learning what 'normal' looks like for your data, helping to set baseline standards across:

- Volume patterns and seasonality
- Freshness expectations and acceptable latency
- Schema stability and structural relationships
- Distribution characteristics and value ranges
- Lineage dependencies and propagation patterns

Trend 2: Cost-Aware Data Tiering and FinOps Integration



The Cost-Quality Connection

Global public cloud spending will reach \$723 billion in 2025, up from \$491 billion in 2022, representing 47% growth over three years.²¹ This explosive growth has put data teams in a difficult position. They are expected to deliver more value from data while also controlling costs. The pressure comes from both directions: business leaders demanding faster insights and AI capabilities, and finance leaders demanding accountability for every dollar spent.

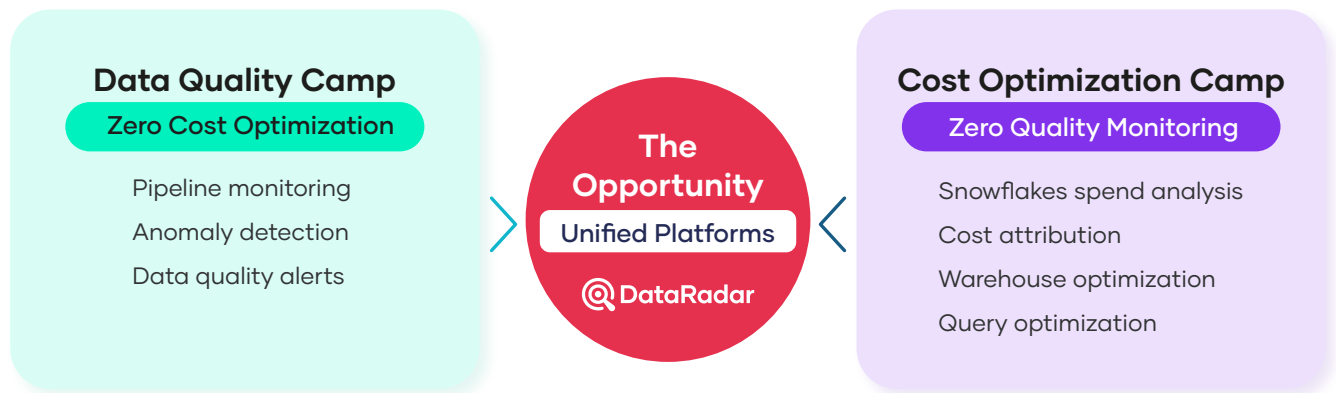
However, the market presents a difficult choice: focus on data quality tools or opt for cost-optimization solutions. This bifurcation forces organizations into an uncomfortable compromise. Choose a data quality platform, and you gain visibility into pipeline health but remain blind to cost drivers. Choose a cost-optimization platform, and you can reduce spend but have no way to tell whether you are cutting costs at the expense of data integrity.

This kind of split leads to tool overload, integration challenges, and overlooked issues. Organizations end up managing multiple vendors, conducting multiple security reviews, and maintaining multiple budgets for capabilities that should be unified. Worse, the lack of integration creates dangerous blind spots. A cost-optimization tool might recommend killing an expensive query without realizing it feeds a critical revenue report. A data-quality tool might flag an anomaly without understanding that the root cause is a misconfigured warehouse that is burning through credits.

The key takeaway is that cost optimization and data quality go hand in hand. You cannot optimize costs intelligently without understanding the quality implications. You cannot ensure quality efficiently without understanding the cost drivers. Organizations that treat these as separate concerns end up overspending on both tools and cloud resources.

The Market Bifurcation

Market Bifurcation



The market forces customers to choose between quality and cost. The opportunity is unified visibility across both

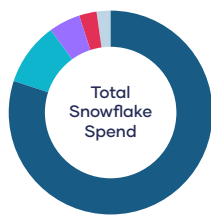
Data Quality Camp	Cost Optimization Camp
Strong observability capabilities	Strong intelligent optimization
ZERO cost optimization features	ZERO data quality monitoring
You still need a separate tool for FinOps	They're blind to quality issues causing costs

Understanding Snowflake Cost Drivers

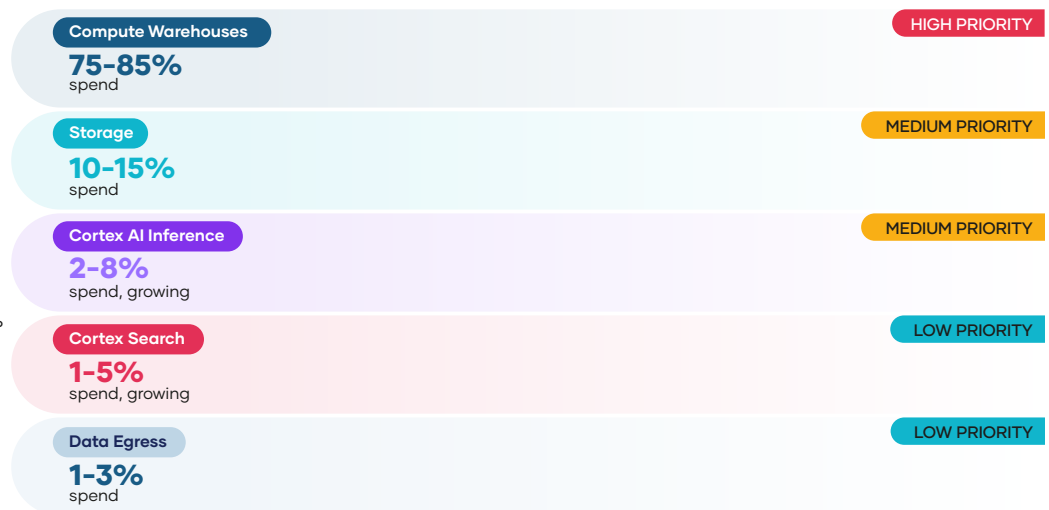
Before optimizing costs, you must understand where they come from. Snowflake's consumption-based pricing model creates multiple cost drivers, each requiring a distinct optimization approach.

Snowflake Cost Breakdown

Typical Enterprise Cost Distribution



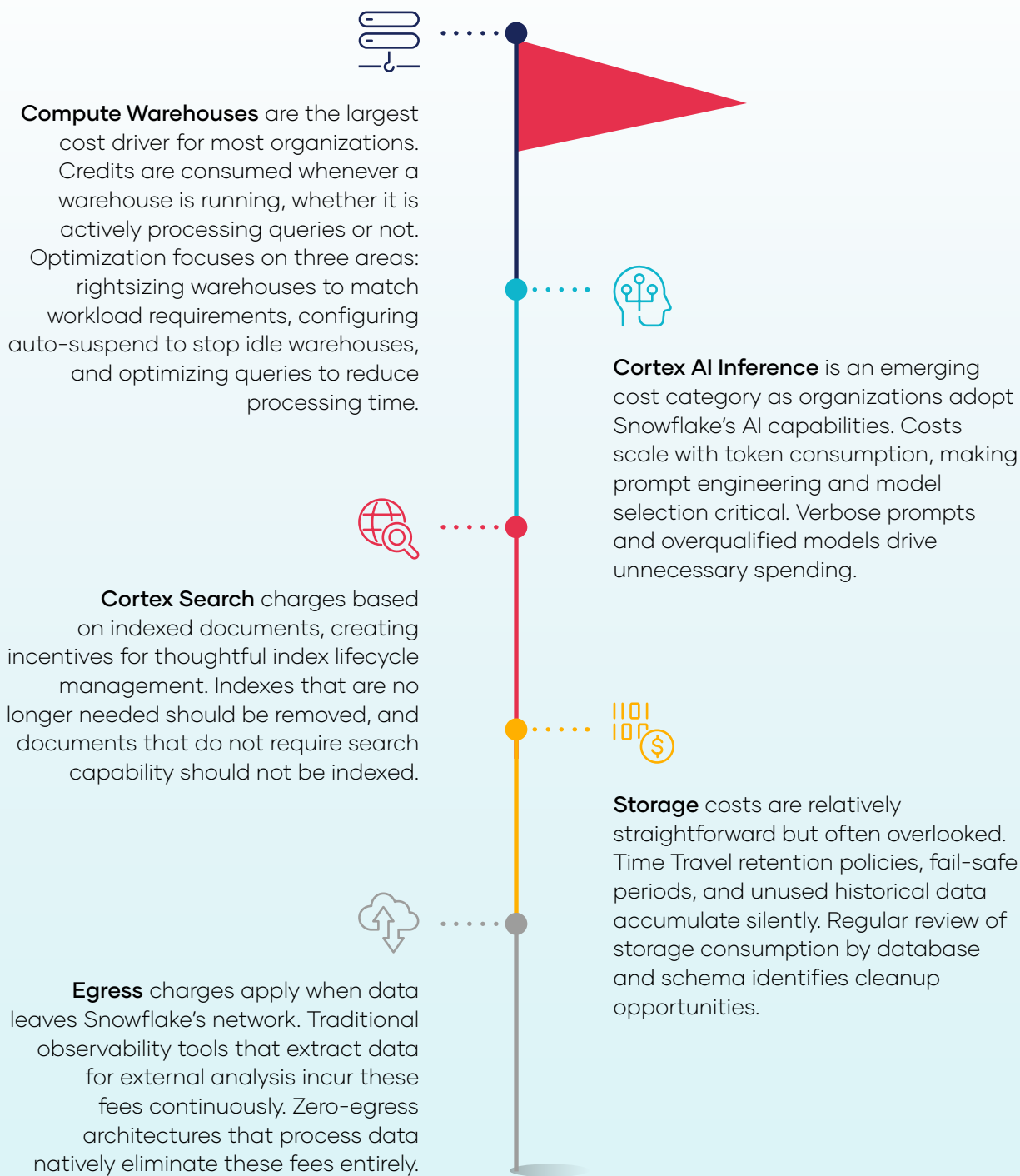
- Compute Warehouses: 80%
- Storage: 10%
- Cortex AI Inference: 5%
- Cortex Search: 3%
- Data Egress: 2%



Optimization Strategy

Compute warehouses account for 75-85% of Snowflake costs, so optimization should focus there. However, managing costs requires visibility across storage, AI inference, and egress, which accumulate over time. Unified platforms help teams monitor, attribute, and optimize all cost areas.

Compute Warehouses	\$2-4 per credit	Rightsizing, auto-suspend, query optimization
Cortex AI Inference	Per million tokens	Prompt optimization, model selection
Cortex Search	Per indexed document	Index lifecycle management
Storage	Per TB/month	Data lifecycle, Time Travel policies
Egress	\$0.09/GB	Zero-egress architecture

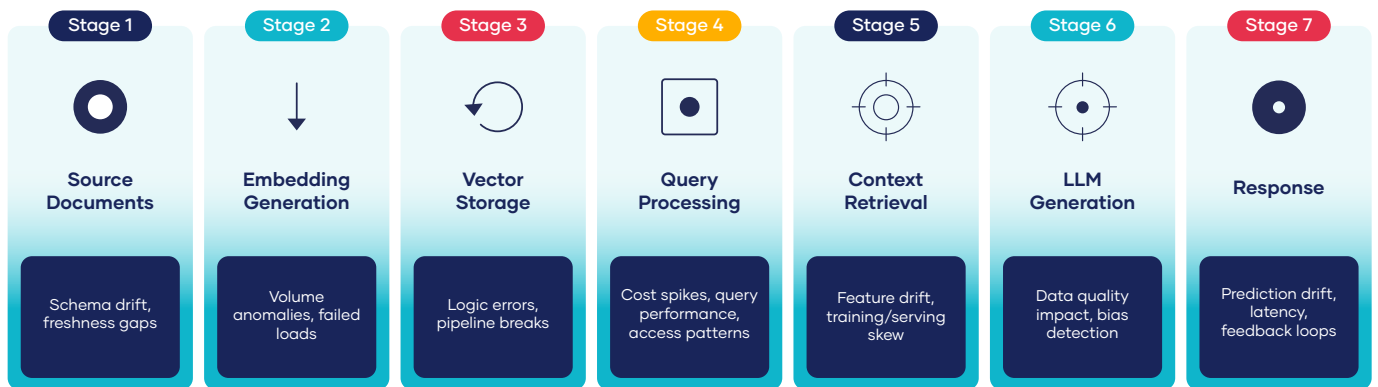


Understanding these cost drivers enables targeted optimization. Rather than applying generic cost-cutting measures, organizations can focus efforts where they will have the greatest impact.

Trend 3: Unified Data and AI Observability

As organizations move from traditional analytics to AI-powered applications, data pipelines have grown increasingly complex. Each stage of the journey from raw data to model output presents distinct observability challenges that legacy monitoring tools were never designed to address.

Data-to-AI Pipeline Flow



Key Insight

AI success depends on data foundation quality. If observability gaps exist in Stages 1-4, every downstream AI application inherits those blind spots.



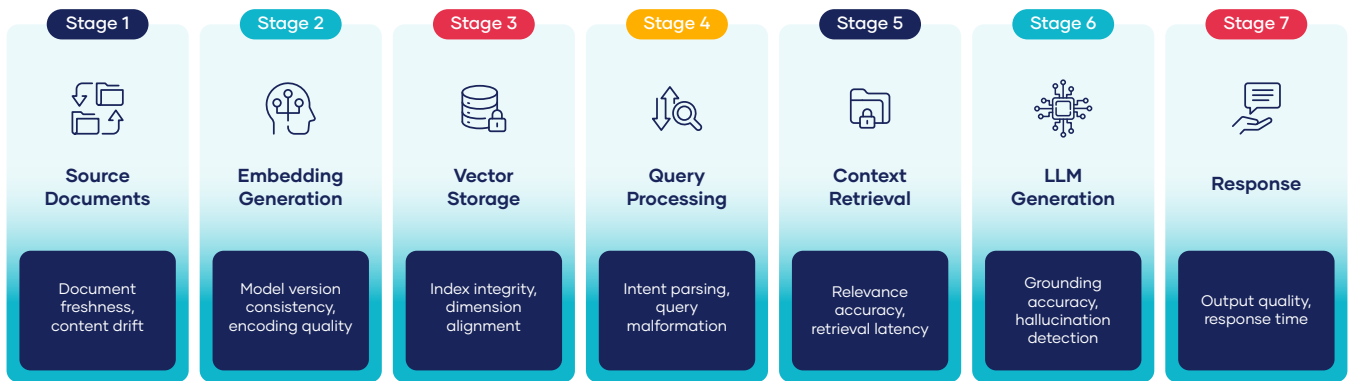
The New Data Stack


Modern AI applications gather data through intricate, multi-step processes that combine classic analytics with machine learning inference. Each step introduces its own quality challenges. Relying solely on traditional observability tools that show only dashboards cannot fully capture this complexity.

RAG Architecture Challenges

Retrieval-Augmented Generation has emerged as the dominant pattern, with the global RAG market valued at \$1.24 billion in 2024 and projected to reach \$1.85 billion in 2025²³

RAG Architecture Diagram





Key Insight
RAG pipelines introduce new observability requirements at every stage. Traditional monitoring tools were not designed for this architecture.

RAG Component	Observability Challenge
Embedding Freshness	Are vector representations current or based on stale source data?
Retrieval Relevance	Is the system returning contextually appropriate documents?
Context Completeness	Does retrieved context provide sufficient information?
Semantic Accuracy	Do generated responses accurately reflect source documents?

Trend 4: Observability for Agentic AI Governance

The shift from assisted AI to autonomous AI marks a fundamental change in how organizations must approach data quality. As AI systems move from making recommendations to taking actions, the consequences of poor data escalate from inconvenient to potentially catastrophic. This new reality demands a governance architecture purpose-built for autonomous operations.

Agentic AI Control Plane

AI Agents: Autonomous systems requesting data and taking actions

Control Plane: The governance layer (four quadrants)

- | | |
|---|--|
| <p>1 Pre-Action Validation
Verify data quality before agent acts</p> | <p>2 Policy Enforcement
Apply guardrails, permissions, business rules</p> |
| <p>3 Action Logging
Complete audit trail of every agent decision</p> | <p>4 Outcome Monitoring
Track results, feed back into quality metrics</p> |

Data Sources: Databases, APIs, systems the agents interact with



Key Insight

Agentic AI requires a governance layer between autonomous agents and your data. The control plane ensures validation, policy enforcement, logging, and monitoring before any action is taken



The Stakes Are Higher

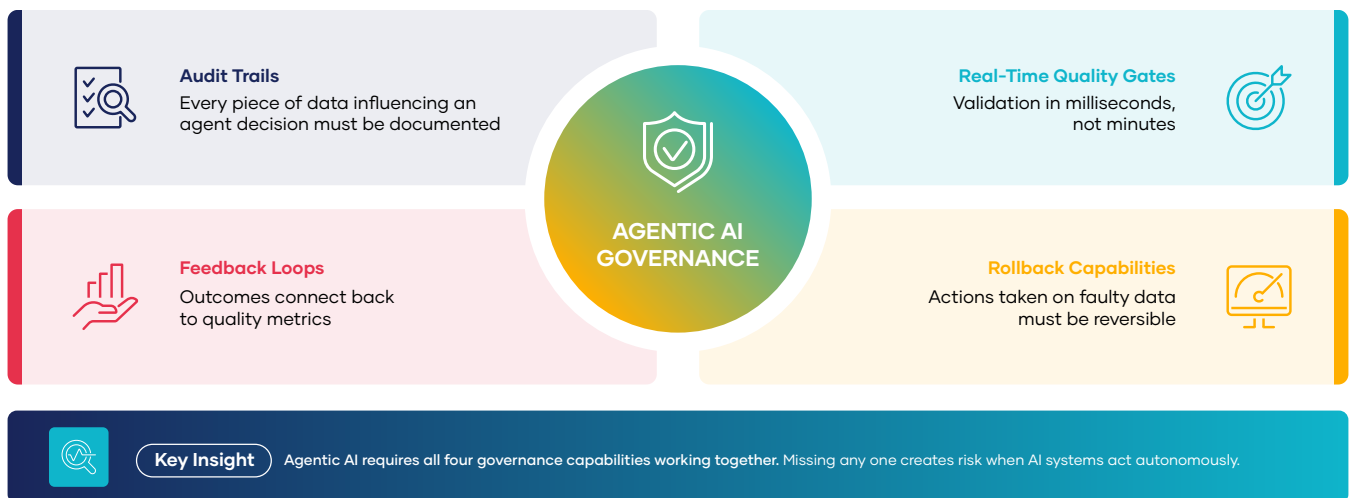
Agentic AI systems take independent actions with minimal human supervision, creating unprecedented demands for high-quality data. When an AI agent performs transactions, updates records, or interacts with customers, the impact of poor data is immediate and potentially irreversible. Unlike traditional analytics, where bad data leads to flawed reports, agentic systems can execute decisions that affect revenue, customer relationships, and regulatory compliance in real time.

Leading analysts projects that by 2028, 33% of enterprise applications will include agentic AI, up from less than 1% in 2024.²⁴ AI enterprise IT market is projected to reach \$182.9 billion by 2034.²⁵ This rapid adoption timeline leaves little time for organizations to develop governance frameworks reactively.

New Governance Requirements

Traditional data privacy governance frameworks assume human review before consequential decisions. Agentic AI challenges this assumption. Organizations must implement proactive AI governance to keep up with evolving AI regulations, while maintaining the rigor of human oversight. This requires a new set of capabilities that work together as an integrated system.

Agentic AI Requirements



Each of these four capabilities addresses a specific failure mode in autonomous systems:

1. Complete audit trails: Every piece of data influencing an agent decision must be documented. When an agent makes an unexpected decision, teams need to trace exactly which data points contributed to that outcome.

2. Real-time quality gates: Validation in milliseconds, not minutes. Agents operate at machine speed, and quality checks must keep pace without introducing latency that defeats the purpose of automation.

3. Feedback loop integration: Outcomes connect back to quality metrics. When agent actions produce suboptimal results, those outcomes must automatically inform upstream data quality scoring.

4. Rollback capabilities: Actions taken on faulty data must be reversible. Organizations need the ability to unwind agent decisions when data quality issues are discovered after the fact.

Organizations that deploy agentic AI without these governance capabilities are essentially allowing autonomous systems to operate without guardrails. The observability layer becomes the mechanism that makes agentic AI safe for enterprise deployment.

Trend 5: Observability as Business Value Driver

From Cost Center to Value Creator

For years, data infrastructure was treated as a cost center: a necessary expense to keep the lights on. Data observability followed the same pattern, justified by incident-reduction metrics that executives struggled to link to business outcomes. Poor data quality costs organizations an average of \$12.9 million per year²⁶, yet many teams struggle to articulate this impact in terms that resonate with leadership.

Data observability has evolved from a technical best practice into a recognized strategic priority — one with measurable impact on data reliability, AI readiness, and operational cost.¹¹ The shift requires translating technical metrics into language executives understand and care about.

Technical to Business Translation

What Data Teams See

MTTD:
4.2 hrs → 1.1 hrs

MTTR:
8 hrs → 2.5 hrs

Incidents/month: 47 → 12
47 → 12

Pipeline failures: 23 → 6
23 → 6

What Executives Hear

\$2.4M
Revenue protected annually

1,200
Engineering hours saved per quarter

\$850K
Compliance penalties avoided

3
Additional AI models in production



Key Insight

Technical metrics only matter when translated into business outcomes. Observability enables this translation

The Translation Gap

When data teams report “MTTD improved from 4.2 hours to 1.1 hours,” executives hear jargon. When they report “revenue was protected because we caught the pricing feed error before it reached customers,” executives hear value.

The metrics are connected: faster detection directly enables revenue protection. But data teams speak in technical terms because that is what they measure; executives speak in business terms because that is what they are accountable for. Organizations that bridge this gap unlock budget, executive sponsorship, and strategic priority for data initiatives.



Four Value Dimensions

Business value from data observability flows through four dimensions:

Value Dimension	Key Questions	Business Impact
Revenue Protection	What revenue depends on accurate, timely data?	Cost of decisions made on faulty information
Operational Efficiency	Time spent on firefighting vs. improvement?	Data team productivity gains
Risk Mitigation	Regulatory, reputational, operational risks?	Compliance penalties, ops costs and litigation avoidance
Innovation Enablement	What becomes possible with trustworthy data?	Opportunity cost of unreliable data



Revenue Protection is often the largest and most tangible category. When a pricing algorithm receives bad data, it makes poor pricing decisions. When a fraud model receives stale data, it misses fraud. Each scenario has a dollar value.



Operational Efficiency directly affects headcount and productivity. Leading analysts report that nearly one-third of analysts spend more than 40% of their time vetting and validating data before it can be used for decision-making²⁷, leaving little capacity for innovation. Observability that reduces this burden frees capacity without adding headcount.



Risk Mitigation gains urgency as regulations tighten. DORA (the Digital Operational Resilience Act), the EU AI Act, and evolving AI and privacy laws and regulations create real financial exposure from data quality failures. Observability provides the documentation and controls regulators expect.

Innovation Enablement is the hardest to quantify but often the most strategic. AI initiatives that stall due to data quality issues incur significant opportunity costs.

Building the Business Case

To shift observability from cost center to value driver, quantify impact across all four dimensions:

1. Identify revenue-critical data flows: Which pipelines, if they fail, directly impact revenue? What is the estimated cost per hour of downtime or bad data?

2. Assess team time allocation: What percentage of data engineering time goes to reactive work? What is the fully-loaded cost of that time?

3. Map litigation and regulatory exposure: Which compliance requirements depend on data quality? What are the potential consequences for violations?

4. Connect to strategic initiatives: Which AI or analytics projects are blocked by data trust issues? What is the projected value of those initiatives?

The resulting business case speaks executive language: revenue protected, costs avoided, risks mitigated, opportunities enabled.

Trend 6: Data Productization and Self-Service Analytics

The centralized data team model is reaching its limits. As organizations scale their data operations, funneling all requests through a single team creates bottlenecks that slow innovation and frustrate business users. A new architectural pattern is emerging that distributes data ownership while maintaining the governance and interoperability enterprises require.

Data Mesh Architecture



The Data Mesh Philosophy

Data mesh principles offer a decentralized approach in which domain teams own their data as products, transforming how organizations view data ownership, quality, and governance. According to IBM research, 82% of enterprises report that data silos disrupt critical workflows, and 68% of enterprise data remains unanalyzed.²⁸

This philosophy addresses a persistent challenge: centralized data teams become bottlenecks. They lack domain expertise to interpret data and determine how it should be used. Data mesh redistributes this responsibility. The sales team owns sales data. The product team owns product data. Each domain understands its data deeply and takes accountability for its quality and availability.

What Makes a Data Product

Not every dataset qualifies as a data product. Data products are intentionally designed assets with specific characteristics:²⁹



- **Discoverable:** Easily found through catalogs and search
- **Addressable:** Uniquely identifiable and accessible
- **Understandable:** Well-documented with clear semantics
- **Trustworthy:** Reliable, accurate, up-to-date with visible quality metrics
- **Self-describing:** Contains metadata describing content and structure
- **Interoperable:** Easily combined with other data products
- **Secure:** Access controlled to protect sensitive information



The Observability Challenge

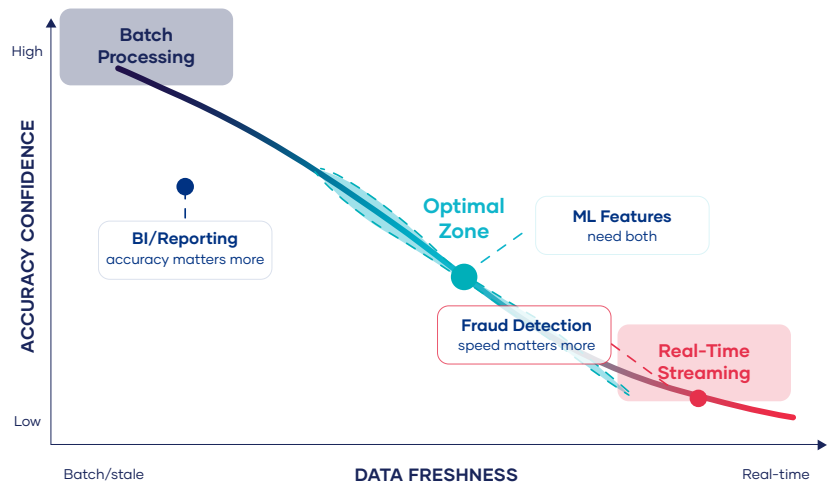
Data mesh addresses the ownership problem but introduces new observability challenges. When data products are distributed across dozens of domains, visibility becomes fragmented. Each domain may monitor its own products, but no one sees the full picture.

Unified observability addresses this challenge. A platform that spans all domains while respecting their autonomy provides the visibility mesh architectures require. Domain teams retain ownership. The platform serves as the connective tissue that makes the mesh function as a coherent whole.

Trend 7: Real-Time Data Quality Monitoring

The shift from batch to streaming has moved from aspiration to expectation. According to Confluent's 2025 Data Streaming Report, 86% of IT leaders now cite data streaming as a strategic or important investment priority.³⁰ Real-time data powers fraud detection, dynamic pricing, operational dashboards, and AI inference pipelines. But speed creates new challenges. When data moves continuously, traditional validation approaches designed for overnight batch jobs no longer apply. Organizations must rethink how they maintain data quality when every millisecond counts.

Freshness vs. Accuracy Tradeoff



Key Insight

Different use cases require different positions on the freshness-accuracy curve. The goal is finding the right balance, not maximizing either dimension.



The Freshness-Accuracy Tradeoff

The shift from batch to streaming is accelerating. Real-time data is now expected, not exceptional. But fresher data creates a paradox: higher velocity increases the risk of acting on incomplete or inaccurate data.

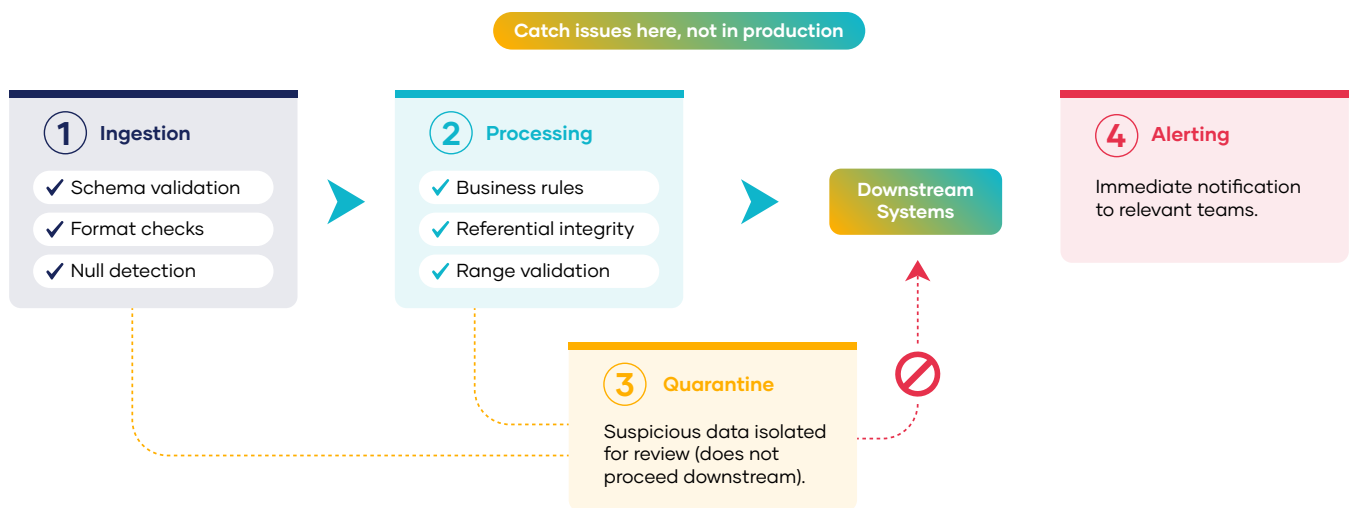
Batch processing allows time for validation, reconciliation, and error correction. The tradeoff is latency. By the time data is validated, it may be hours old. Streaming flips this equation. Data arrives continuously, and consumers expect immediate access. Checks must happen inline, in milliseconds, or not at all.

The answer is not choosing one extreme over the other. It is understanding where each use case falls on the curve and building validation appropriate to that position. BI dashboards can tolerate latency in exchange for accuracy. Fraud detection cannot. ML feature pipelines need both, placing them in the optimal zone where tradeoffs must be carefully managed.

Shift-Left Quality Monitoring

Traditional data quality approaches treat validation as a downstream activity. Data flows through pipelines and lands in warehouses before quality checks reveal problems. By then, bad data has already propagated to dashboards, reports, and ML models. Remediation becomes expensive and disruptive. A shift-left approach moves validation upstream, catching issues at ingestion before they can spread. The result is faster detection, lower remediation costs, and greater confidence in downstream outputs.

Shift-Left Pipeline



Key Insight

The earlier you catch issues, the cheaper they are to fix. Shift-left means validation happens at the start, not the end.



The Shift-Left Philosophy

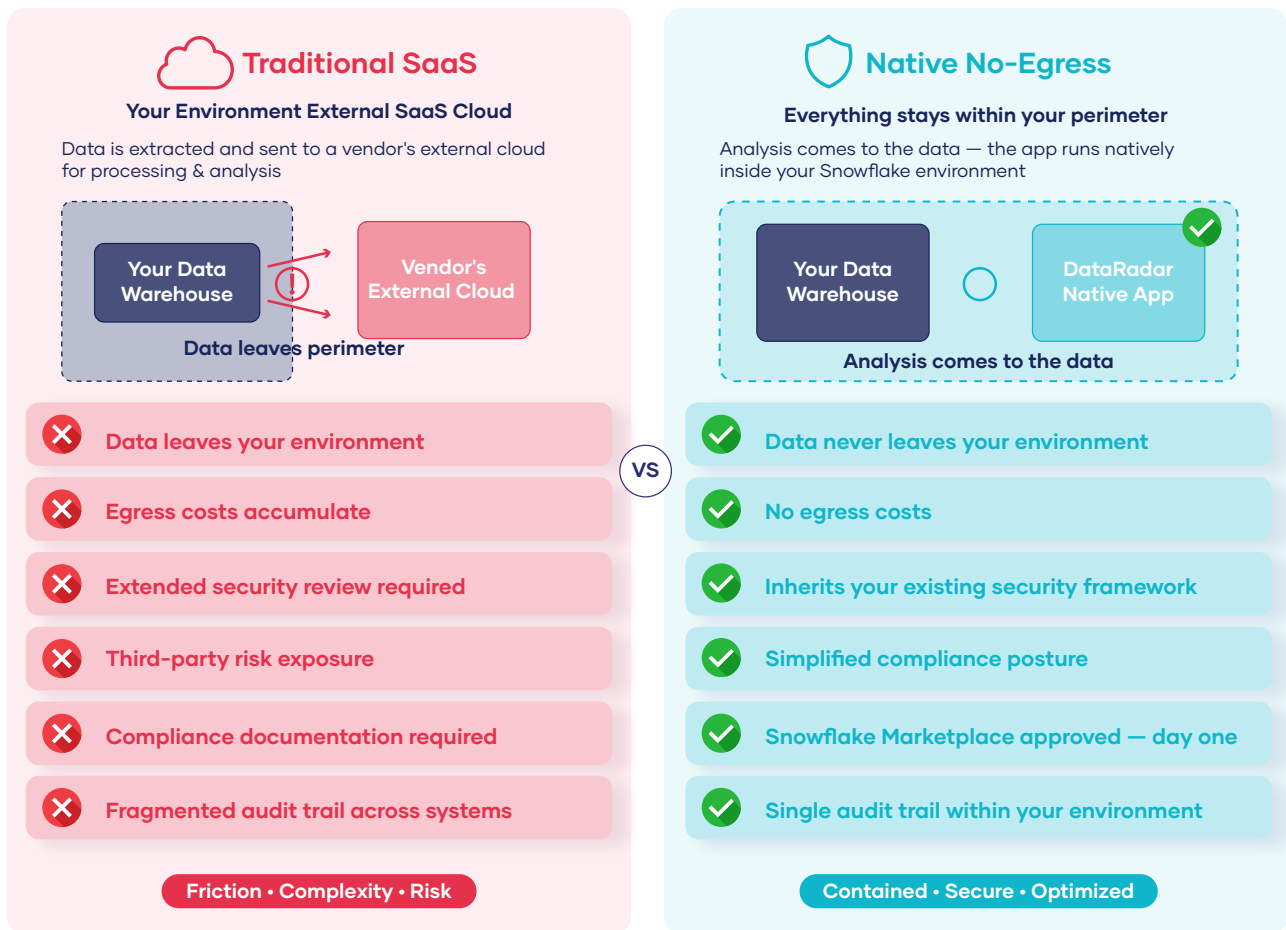
The term “shift-left” comes from software development, where catching bugs early costs far less than fixing them in production. The same principle applies to data.

A schema violation detected during ingestion is a minor inconvenience. The same violation discovered after data has propagated through downstream systems, been fed into ML models, and populated dashboards becomes a major incident requiring significant remediation.

Trend 8: Cloud-Native Observability Architectures

As organizations embrace multicloud environments and cloud-native architectures, the complexity of their technology stacks has grown significantly. According to Dynatrace's 2024 State of Observability report, the average multicloud environment now spans 12 platforms and services, and organizations use an average of 10 monitoring and observability tools to manage applications, infrastructure, and user experience.³¹ This fragmentation raises a critical question: where should observability processing occur? The traditional answer has been external SaaS platforms that pull data from your environment. But a new architectural pattern is emerging that keeps data within your security perimeter while still delivering comprehensive observability.

Zero-Egress Comparison



Key Insight

Where your data goes matters. No-egress architecture keeps data inside your perimeter while still delivering full observability.



Key Insight
Where your data goes matters. No-egress architecture keeps data inside your perimeter while still delivering full observability.

The Zero-Egress Imperative

One of the most critical architectural decisions organizations face is whether their observability tools should run within or outside the data platform.

Traditional SaaS observability tools require data extraction. They pull data, or metadata about your data, into their cloud environment for analysis. This approach made sense when on-premises infrastructure was the norm and cloud-based tools offered capabilities that were not available internally.

The landscape has changed. Modern data platforms like Snowflake provide the compute power and extensibility to run sophisticated analysis natively. The question is no longer whether cloud-based analysis is possible but whether external analysis is necessary.

For many organizations, the answer is no. And the implications for that answer are significant.



Problem	Impact
Security Risk	Are vector representations current or based on stale source data?
Compliance Issues	Is the system returning contextually appropriate documents?
Egress Costs	Does retrieved context provide sufficient information?
Performance Overhead	Do generated responses accurately reflect source documents?
Dual Security Reviews	Your team AND the vendor's infrastructure

Each row in that table represents a friction point. Security teams must evaluate external platforms. Compliance teams must document data flows. Finance teams must budget for egress fees. Architecture teams must account for latency. The cumulative burden slows adoption and increases the total cost of ownership.

Zero-egress architecture eliminates these concerns by bringing analysis to the data rather than bringing the data to the analysis.



The Native Advantage

Native platforms that run inside your existing data environment inherit your existing security controls:

1. Your existing Role-Based Access Controls (RBAC): No need to replicate permission structures in an external system. The access policies you have already defined apply automatically.

2. Your encryption standards: Data remains encrypted according to your policies. No additional encryption configuration required for a third-party platform.

3. Your network policies: No firewall rules to modify. No VPN tunnels to establish. No external endpoints to secure.

4. Your compliance certifications: If your data platform is already certified for SOC 2, HIPAA, or other standards, native tools inherit that compliance posture.

This inheritance model dramatically simplifies procurement. Security reviews that take months for external platforms can be completed in days for native applications. Budget approvals that require new line items for external vendors can leverage existing platform commitments.

The result is faster time to value. Organizations that would spend quarters evaluating and implementing external tools can deploy native observability in weeks.

Trend 9: Regulatory Pressure and Compliance Automation

The Regulatory Landscape

2024 ... ● ... **2024** saw an intensification of GDPR (the General Data Protection Regulation) enforcement across the European Union (EU), with regulators issuing record fines and expanding their interpretation of data quality obligations. In the United States, CCPA (the California Consumer Privacy Act) amendments strengthened consumer data **rights and introduced new requirements for data accuracy**.

January 2025 ... ● ... In January 2025, the DORA (Digital Operational Resilience Act) took effect, requiring European financial institutions to demonstrate ICT (Information and Communication Technology) operational resilience. Financial services organizations must now prove they can detect, respond to, and recover from data-related incidents through documented processes and measurable metrics.

August 2026 ... ● ... In August 2026, the EU AI Act's high-risk AI requirements will take effect, mandating thorough data lineage and quality logging for systems that affect employment, credit, healthcare, and other sensitive domains.³⁴ Organizations deploying high-risk AI will need to demonstrate that training data meet quality standards and that consequential decisions can be traced back to their data inputs.

2027 ... ● ... **2027** will bring full enforcement of the EU AI Act, along with anticipated amendments to the GDPR, CCPA, and sector-specific regulations and state laws. The regulatory environment will be materially more demanding than it is today.

These are not isolated regulations. They are converging. Data quality, lineage, and governance requirements appear across all of them. Organizations that address these capabilities once, through unified observability, satisfy multiple frameworks simultaneously. Organizations that address them piecemeal face redundant work and compliance gaps.

Regulatory Requirements Matrix

Regulations	Data Quality Monitoring	Data Lineage	Access Controls	Anomaly Detection	Audit Trail	Incident Response
DORA (Digital Operational Resilience Act)	✓	✓	✓	✓	✓	✓
EU AI Act (Effective 2024-2026)	✓	✓		✓	✓	
GDPR (General Data Protection Regulation)	✓	✓	✓		✓	✓
CCPA (California Consumer Privacy Act)	✓	✓	✓		✓	
HIPAA (Health Insurance Portability and Accountability)	✓		✓	✓	✓	✓
SOX (Sarbanes-Oxley Act)	✓	✓	✓		✓	



Key Insight

One observability platform, multiple compliance frameworks addressed. Data quality and lineage appear across nearly every regulation.

Regulation Requirement at a Glance

Regulation	Scope	Observability Requirement
DORA	EU financial institutions and ICT third-party providers	ICT operational resilience, incident reporting, and continuous monitoring
EU AI Act	High-risk AI system use in EU	Data lineage, quality logging, bias monitoring
GDPR/CCPA	Personal data processing	Data access tracking, retention enforcement, and breach detection
HIPAA	US protected health information ("PHI")	PHI access monitoring, audit trails, breach detection
SOX	US public companies	Financial data integrity, change tracking, and audit documentation

Executive Summary: Chapter 2



The Core Insight:

Nine interconnected trends are transforming how enterprise data is managed in 2026. These aren't merely separate forces; they're part of a core transformation that calls for a comprehensive, big-picture approach.



The Bifurcation Problem:

The market has bifurcated into quality-focused and cost-focused tools, compelling organizations to oversee multiple vendors to address interconnected issues.



Market Reality:

The data observability market has grown substantially, reaching a valuation of \$3.15 billion in 2025 and is projected to expand to \$5.45 billion by 2030. Notably, by 2027, approximately 70% of organizations are expected to adopt observability tools, underscoring their growing significance.



The Unified Approach:

Organizations succeeding in 2026 don't treat these trends as separate initiatives—they recognize them as inseparable capabilities that require integrated platforms.





Questions to Ask Your Organization

Strategic


1. Are we treating data observability and cost optimization as separate disciplines, or as one integrated system?
2. Which of the nine trends represents our biggest gap compared to our AI ambitions?
3. How many separate vendors do we currently use for data quality, cost optimization, and governance?
4. What is our strategy for the coming wave of agentic AI systems that take autonomous actions?

Tactical

1. Do we have agentic AI systems that take autonomous actions? If so, how are we ensuring data quality for their inputs?
2. What is our strategy for monitoring RAG architectures and assessing vector database quality?
3. How are we balancing data freshness against validation confidence in streaming pipelines?
4. Which existing and upcoming privacy and AI regulations (e.g., DORA, EU AI Act, GDPR, HIPAA, CCPA state consumer privacy updates, laws, and AI laws) apply to us, and are we prepared?

Financial

1. Can we quantify the business value of our data observability investments in a way executives understand?
2. What percentage of our cloud data spend goes to reprocessing failures versus creating new value?
3. How many 'zombie pipelines' are consuming compute without delivering business value?
4. Does our observability architecture support zero-egress operation, or are we paying egress fees?

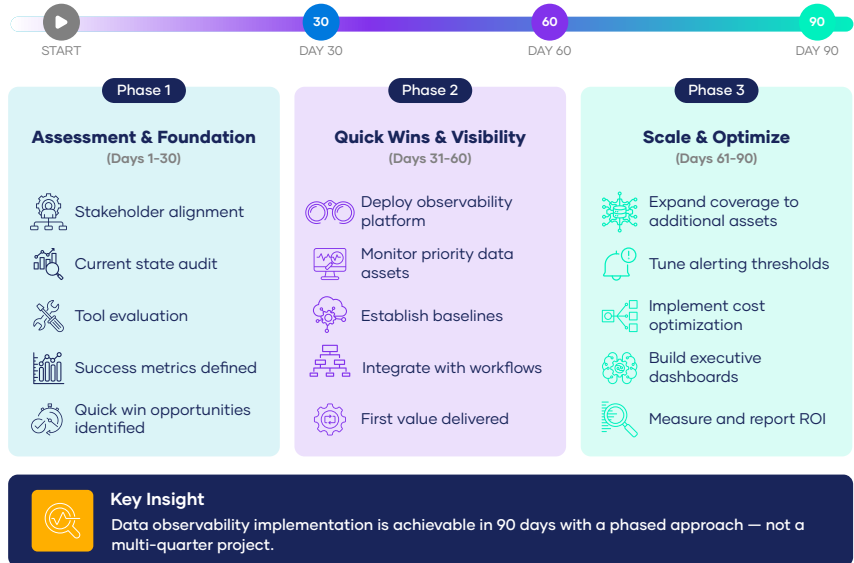


Chapter 3: The 90-Day Implementation Roadmap

Overview

This chapter presents a 90-day roadmap for advancing data infrastructure from an aspirational state to full AI readiness. The framework is practical, achievable, and grounded in patterns observed in successful implementations. According to industry research, organizations that follow structured implementation approaches achieve significantly better outcomes than those that adopt ad hoc approaches.³⁵

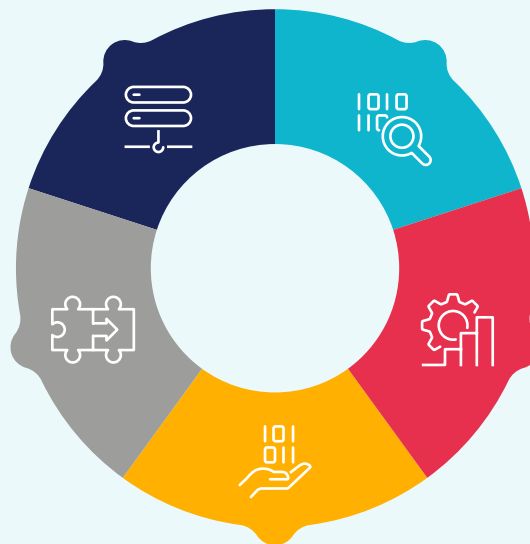
90-Day Timeline Overview



The AI-Ready Data Maturity Model

Assessing your current state requires evaluating five dimensions:

- 1. Data Quality Foundation:** Accuracy, completeness, freshness, and consistency of data assets
- 5. Culture & Skills:** Team capabilities, cross-functional collaboration, and data literacy



- 2. Data Observability:** Visibility into data health, pipeline status, and anomaly detection
- 3. Data Governance:** Policies, ownership, access controls, and compliance documentation

- 4. Architecture & Infrastructure:** Platform capabilities, scalability, and deployment model

Maturity Levels

Organizations progress through five maturity levels across these dimensions:

Level	Name	Description
1	Ad Hoc / Reactive	Informal processes, issues discovered after impact
2	Defined / Repeatable	Basic standards exist, and manual processes
3	Managed / Standardized	Automated processes, comprehensive monitoring
4	Optimized / Proactive	Intelligent capabilities, predictive analytics
5	AI-Ready / Autonomous	Self-healing systems, autonomous operations



Most organizations currently operate at Levels 2-3. Research indicates that data silos and fragmented tooling are the primary barriers to advancement, with 82% of enterprises reporting significant challenges from disconnected data systems.³⁶ The gap between leaders (Levels 4-5) and laggards is widening as AI capabilities become essential to competitive advantage.

Phase 1: Assessment & Foundation (Days 1-30)

The first phase establishes the foundation for everything that follows. Rushing past assessment leads to misaligned priorities and wasted effort.

Week 1-2:

Data Estate Discovery

Begin with a comprehensive inventory of your data landscape:

1. Inventory all data sources, pipelines, and consumers
2. Document current data quality issues and their business impact
3. Map existing monitoring and alerting capabilities
4. Identify AI use cases and their specific data requirements

This discovery work often reveals surprises. Organizations typically underestimate the complexity of their data estate and overestimate their current monitoring coverage.

Week 3-4:

Maturity Assessment & Prioritization

With complete discovery, assess your maturity and set priorities:

1. Complete an AI-Ready Data Maturity assessment across all five dimensions
2. Identify highest-priority gaps based on AI use case requirements
3. Define success metrics and establish baselines
4. Secure executive sponsorship and budget approval

The output of Phase 1 is clarity: you know where you stand, where you need to go, and what success looks like.

Phase 2: Quick Wins & Visibility (Days 31-60)

Phase 2 is about deploying capabilities and demonstrating value quickly. Momentum matters. Teams that deliver early wins build organizational support for broader investment. Teams that get bogged down in procurement, security reviews, or integration challenges lose momentum and risk having their initiatives deprioritized.

The goal of Phase 2 is not perfection. It is visibility. You want to establish baseline monitoring, identify your initial set of actionable insights, and demonstrate to stakeholders that the investment is paying off. Perfection comes later. Momentum comes now.

This phase typically includes deploying your observability platform, connecting critical data sources, setting initial alerting thresholds, and generating your first set of quality and cost reports. By day 60, you should have concrete metrics to share with leadership: incidents detected, costs identified, and time saved.

Why Native Architecture Matters

The deployment model you choose significantly affects time-to-value. Traditional SaaS observability platforms require data extraction, security reviews, network configuration, and often months of procurement. A native architecture eliminates these barriers.

The difference is not marginal. It is the difference between being operational in days versus in quarters. For organizations under competitive pressure or regulatory deadlines, this acceleration can be the difference between making the deadline and missing it.

Native architecture offers four critical advantages:

Four Pillars: Unified, Native, Rapid, Approved

<p>1</p>  <p>Unified</p> <p>Snowflake Native App combining data quality monitoring + cost optimization. One tool replaces two vendors. One security review instead of two.</p>	<p>2</p>  <p>Native</p> <p>Data never leaves your Snowflake environment. No infrastructure provisioning. No data egress. No metadata extraction. Your security team reviews one architecture, not two external SaaS platforms.</p>	<p>3</p>  <p>Rapid</p> <p>Deploy in under 30 minutes from Snowflake Marketplace without help, calls, or proof of concept. Download your 30-day trial and start monitoring and optimizing today, not next quarter.</p>	<p>4</p>  <p>Approved</p> <p>Purchase using existing Snowflake credits. Finance has already approved the expenditure. No new vendor onboarding, no new budget cycle, and no procurement delays.</p>
--	--	--	---



Key Insight

Native architecture delivers four compounding advantages that accelerate time-to-value and reduce deployment friction.



These advantages compound. Native deployment in 30 minutes versus an external deployment in 30 days means you reach Phase 3 while competitors are still in procurement. For organizations subject to DORA or EU AI Act requirements, this acceleration is particularly critical given compliance deadlines.³⁷

Phase 3: Scale & Optimize (Days 61-90)

With a foundation in place and quick wins delivered, Phase 3 focuses on expanding coverage and measuring results.

Week 9-10:

Expand Coverage

Build on initial success by broadening scope:

1. Extend observability to additional data domains
2. Implement cost monitoring and optimization recommendations
3. Integrate with data catalog and governance tools
4. Train data teams on new tools and processes

Week 11-12:

Measure & Plan

The final weeks focus on demonstrating value and planning next steps:

1. Measure progress against baseline metrics established in Phase 1
2. Calculate ROI from quality improvements and cost savings
3. Document lessons learned and best practices
4. Create a roadmap for the next 90 days

Success Metrics

Track progress against these benchmarks, which reflect outcomes observed across successful implementations:³⁸

Metric	Baseline	30-Day Target	90-Day Target
Data quality incidents/month	Measure current	-25%	-50%
Mean time to detection	Measure current	-40%	-60%
Critical pipeline coverage	Measure current	-25%	-40%
Critical pipeline coverage	0%	50%	80%+
Cloud cost savings	\$0	10-15%	20-30%
Zombie pipelines eliminated	Unknown	50%	100%

These targets are achievable. Organizations following this roadmap consistently hit or exceed them.

Executive Summary: Chapter 3



The Core Framework:

The AI-Ready Data Maturity Model assesses readiness across five dimensions: Data Quality Foundation, Data Observability, Data Governance, Architecture & Infrastructure, and Culture & Skills.



The 90-Day Roadmap:

A practical implementation plan that moves from Assessment (Days 1-30) through Quick Wins (Days 31-60) to Scale & Optimize (Days 61-90).



Current State:

Most organizations operate at Levels 2-3 (Defined/Managed) across dimensions. The gap between leaders (Levels 4-5) and laggards is widening.



Expected Outcomes:

Organizations following this roadmap typically achieve a 50% reduction in data quality incidents, 60% faster detection, and 20-30% savings in cloud costs within 90 days.



Questions to Ask Your Organization

Before beginning your 90-day implementation, engage stakeholders with these questions:

Strategic

1. What is our current maturity level across each of the five dimensions?
2. Which AI use cases are the highest priority, and what are their specific data requirements?
3. Who is the executive sponsor?

Tactical

1. Do we have baseline metrics for data quality incidents, detection time, and resolution time?
2. Can we identify 10-20 critical data assets to prioritize in Phase 2?
3. What compliance deadlines are we facing?

Financial

1. Is the budget already approved, or can we use existing Snowflake credits?
2. How many tools are we currently using for data quality and cost optimization?

Chapter 4: The Agentic AI Era—Future-Proofing for 2027 and Beyond

Organizations that develop adaptable, AI-ready data infrastructures today will secure value irrespective of the specific technologies that emerge in the future.

The Agentic AI Revolution

By 2028, it is projected that 33% of enterprise applications will incorporate agentic artificial intelligence.³⁸ These systems go beyond mere data analysis; they take autonomous actions such as scheduling meetings, executing trades, adjusting inventory levels, and responding to customer inquiries without human intervention. This transition from advisory AI to autonomous AI necessitates new requirements for data observability.

Agentic AI Requirement



Action Audit Trails

Complete traceability of the data that informed every agent’s decision is required. When an agent takes action, it must be possible to precisely identify which data inputs influenced that decision. This requirement is mandatory and constitutes a governance obligation.

Real-Time Quality Gates

Data validation occurs within microseconds, as agents operate at machine speed. Quality assurance processes that require minutes or hours are ineffective. Validation must be conducted in real time, at the point of data consumption.

Rollback Capabilities

Ability to reverse actions taken on insufficient data. When an agent acts on incorrect data, you need the ability to undo the damage. This requires not only detection but also automated remediation workflows.

Feedback Loop Integration

Agent outcomes automatically feed back into quality metrics. Did the action succeed or fail? The outcome should inform data quality scoring, creating a continuous improvement cycle.

Organizations deploying agentic AI without these safeguards are building systems that will fail in unpredictable and potentially catastrophic ways.³⁹

Autonomous Data Operations

The evolution of DataOps reflects the broader shift toward autonomy. Understanding this progression helps organizations plan their maturity journey.

Era	Capability	Characteristics
2020	Manual Operations	Reactive troubleshooting, tribal knowledge, and firefighting culture
2022	Automated Monitoring	Dashboards, scheduled checks, and alert-based response
2024	Intelligent Alerting	ML-powered anomaly detection, prioritized alerts, reduced noise
2026	Self-Healing Systems	Automated remediation, predictive maintenance, minimal human intervention
2028	Autonomous Operations	AI-driven optimization, continuous adaptation, human oversight only

Most organizations currently function within the spectrum of automated monitoring and intelligent alerting. Leadership has begun implementing self-healing capabilities. Leading analysts predict that by 2028, at least 15% of day-to-day work decisions will be made autonomously by agentic AI — up from zero percent in 2024 — and one in three enterprise software applications will embed agentic capabilities (Gartner, 2025). For data teams, this makes autonomous operations not a future consideration, but an imminent operational reality.



The pivotal insight is that each phase is contingent upon the preceding one. It is not permissible to advance directly to autonomous operations without first establishing the essential foundations of automated monitoring, intelligent alerting, and self-healing mechanisms. The 90-day roadmap outlined in Chapter 3 provides the necessary groundwork for this progression.

The Multi-Modal Data Challenge

By 2027, the adoption of multi-modal AI systems — integrating text, images, audio, and structured data — is expected to reach 40% of all generative AI solutions, up from just 1% in 2023 (Gartner, 2023). For example, a customer service AI could simultaneously analyze a customer's tone of voice, facial expressions, purchase history, and written complaints. Similarly, manufacturing AI may combine sensor readings, visual inspection outcomes, maintenance logs, and supply chain data.

This creates new data quality challenges that traditional observability tools were not designed to address:

- 1. Cross-Modal Consistency:** Do the audio and text sentiments align? Do the visual inspection results align with sensor readings? Inconsistencies across modalities indicate data quality issues, but detecting them requires understanding relationships between fundamentally different data types.
- 2. Temporal Synchronization:** Multi-modal systems require data from different sources to be synchronized in time. A video frame from 10 seconds ago, combined with sensor data from the present, produces unreliable results. Observability must track temporal alignment across modalities.
- 3. Quality Metrics for Unstructured Data:** What does 'data quality' mean for an image or an audio file? Traditional metrics such as completeness and accuracy need new definitions for unstructured data types.



Organizations investing in data observability today should ensure their platforms can evolve to address multi-modal challenges or plan for platform transitions as these requirements arise.

Principles That Endure

While technologies evolve rapidly, certain principles remain constant.

These five principles will be as relevant in 2030 as they are today:



1. Data Quality Is Foundational

Every advancement in artificial intelligence highlights the significance of reliable data. More sophisticated models do not mitigate the impact of poor data; instead, they exacerbate its effects. As AI capabilities expand, the consequences of inadequate data quality increase correspondingly.

2. Observability Enables Innovation

Visibility facilitates experimentation; blind spots lead to failure. Teams capable of perceiving their data systems' states can iterate rapidly, identify issues proactively, and foster confidence in innovative strategies. Conversely, teams operating without insight tend to proceed slowly and experience frequent failures.

3. Governance Must Scale

As the proliferation of artificial intelligence systems persists, it is essential to implement governance mechanisms through automation. Manual governance procedures, which may be effective for 10 data products, become impractical when managing 1,000. The only feasible approach is to integrate automated governance into data pipelines.

4. Architecture Matters

The architectural decisions made today either constrain or facilitate future capabilities. Native deployment models that prevent data movement, unified platforms that eliminate silos, and adaptable architectures that support new data types are among the decisions that accrue over time.

5. Cost and Quality Are Unified

Effective cost optimization is unattainable without a comprehensive understanding of quality. Conversely, trust in quality cannot be established without awareness of associated costs. The artificial segmentation of these two considerations engenders oversight. The future is poised to favor platforms that effectively address both aspects.

Executive Summary: Chapter 4



The Core Message:

Technologies will evolve, yet the fundamental principles will remain unchanged. Organizations that establish adaptable, AI-ready data foundations today will be able to capture value, regardless of which specific AI capabilities develop in the future.



Autonomous Data Operations:

The progression from manual operations to automated monitoring, intelligent alerting, and self-healing systems, culminating in fully autonomous operations by 2028. Each phase builds on its predecessor.



Enduring Principles:

Five principles that remain relevant despite technological change: data quality is foundational, observability enables innovation, governance must scale, architecture matters, and cost and quality are aligned.



Agentic AI Revolution:

By 2028, 33% of enterprise applications will include agentic AI. These autonomous systems will require new governance controls, including audit trails, real-time quality gates, rollback capabilities, and integration with feedback loops.



Multi-Modal Challenge:

2027 will see the proliferation of multimodal AI systems that combine text, images, audio, and structured data.⁴⁰ This shift will create new quality challenges, including cross-modal consistency, temporal synchronization, and metrics for unstructured data.





Questions to Ask Your Organization

As you plan for 2027 and beyond, engage stakeholders with these questions:

Strategic

1. Are we planning for agentic AI, or will it catch us unprepared?
2. Where are we on the DataOps maturity curve?
3. Do our architectural choices today enable or constrain tomorrow's capabilities?

Tactical

1. Can our current observability platform evolve to handle multi-modal data?
2. Do we have audit-trail capabilities sufficient for autonomous AI governance?
3. What is our plan for real-time quality gates?

Financial

1. Are we investing in platforms that will remain relevant, or will we need to re-platform?
2. What is the cost of being unprepared for agentic AI?



Chapter Resources

90-Day Data Observability Readiness Checklist

Use this checklist to track your progress through each phase. Check off items as you complete them.

Phase 1

Assessment & Foundation (Days 1-30)

- Identified executive sponsor and secured commitment
- Completed inventory of all data sources, pipelines, and consumers
- Documented current data quality issues and business impact
- Mapped existing monitoring and alerting capabilities
- Identified priority AI use cases and their data requirements
- Completed AI-Ready Data Maturity assessment (all five dimensions)
- Established baseline metrics (incidents, MTTD, MTTR)
- Defined success criteria and 90-day targets
- Identified 10-20 critical data assets to prioritize
- Secured budget approval or confirmed Snowflake credit availability

Phase 2

Quick Wins & Visibility (Days 31-60)

- Selected and deployed a data observability platform
- Connected platform to priority data sources
- Established baselines for priority data assets
- Configured quality rules for critical pipelines
- Integrated alerts with team workflows (Slack, Teams, ServiceNow)
- Tuned alerting thresholds to reduce noise
- Achieved 50% coverage of critical pipelines
- Delivered first quick win (documented and communicated)
- Trained core team on platform usage
- Reviewed progress against 30-day targets

Phase 3

Scale & Optimize (Days 61-90)

- Extended observability to additional data domains
- Achieved 80%+ coverage of critical pipelines
- Enabled cost monitoring and attribution
- Implemented cost optimization recommendations
- Integrated with data catalog and governance tools
- Trained broader team on tools and processes
- Identified and eliminated zombie pipelines
- Measured progress against baseline metrics
- Calculated and documented ROI
- Created roadmap for next 90 days

Download this checklist at: dataradar.io/90-day-checklist



Conclusion: Building the Foundation for AI Success

The gap between AI ambition and execution almost always comes down to the data foundation. Organizations that invest in observable, trustworthy, AI-ready data operations will successfully deploy AI at scale. Those that don't will continue to struggle with pilots that never reach production.

The nine trends outlined in this guide represent both the challenges you'll face and the opportunities you can seize. Success requires treating these trends not as isolated issues but as interconnected parts of a unified system.

The Path Forward

- Assess your current maturity across all five dimensions
- Prioritize gaps based on your specific AI use cases
- Implement using the 90-day roadmap as your guide
- Unify your approach, don't settle for fragmented tools
- Measure progress against clear, business-aligned metrics
- Iterate continuously as technologies and requirements evolve

Trust Your Data.

Control Your Costs.

Power Your AI.

Organizations that succeed in 2026 and beyond don't treat data observability, quality, cost optimization, and governance as separate initiatives. They recognize these capabilities as inseparable because, in practice, they are.



The winning approach:

Unified platforms that eliminate false choices. Native architectures that simplify security and accelerate deployment. Integrated visibility shows you both what's wrong with your data and what it's costing you.

The time to act is now. The cost of inaction—failed AI initiatives, wasted investments, and missed opportunities—continues to grow daily.

Your AI is only as good as your data. Build the trusted foundation for AI success.

Appendix A:

Complete Source Citations

All statistics and research findings in this playbook are cited below with links to sources where available.

¹ Nash, K. S. (2025, March 25). 88% of AI pilots fail to reach production — but that's not all on IT. CIO. <https://www.cio.com/article/3850763/88-of-ai-pilots-fail-to-reach-production-but-thats-not-all-on-it.html>

² Gartner. (2021). The cost of poor data quality. <https://www.gartner.com/en/data-analytics/topics/data-quality>

³ Precisely & Drexel University LeBow College of Business. (2024). 2025 outlook: Data integrity trends and insights. <https://www.precisely.com/press-release/new-global-research-points-to-lack-of-data-quality-and-governance-as-major-obstacles-to-ai-readiness>

⁴ Gartner. (2021). The cost of poor data quality. <https://www.gartner.com/en/data-analytics/topics/data-quality>

⁵ International Data Corporation. (2024). AI pilot-to-production research [Research report in partnership with Lenovo]. As cited in Nash, K. S. (2025, March 25). 88% of AI pilots fail to reach production. CIO. <https://www.cio.com/article/3850763/88-of-ai-pilots-fail-to-reach-production-but-thats-not-all-on-it.html>

⁶ Gartner. (2023). How to stop data quality undermining your business. <https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business>

⁷ Precisely & Drexel University LeBow College of Business. (2024). 2025 outlook: Data integrity trends and insights. <https://www.precisely.com/press-release/new-global-research-points-to-lack-of-data-quality-and-governance-as-major-obstacles-to-ai-readiness>

⁸ Forrester Research. (2018). Data performance management is essential to demonstrate data ROI. <https://www.forrester.com/report/Data-Performance-Management-Is-Essential-To-Prove-Datas-ROI/RES83344>

⁹ Redman, T. C. (2017). Seizing the opportunity in data quality. MIT Sloan Management Review. <https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/>

¹⁰ International Data Corporation. (2024). AI pilot-to-production research [Research report in partnership with Lenovo]. As cited in Nash, K. S. (2025, March 25). CIO. <https://www.cio.com/article/3850763/88-of-ai-pilots-fail-to-reach-production-but-thats-not-all-on-it.html>

¹¹ Gartner, Inc. (2026, February 23). Market guide for data observability tools. https://www.linkedin.com/posts/monarakibe_gartner-just-announced-the-market-guide-for-activity-7432421851848679424-pQ5K#:~:text=Gartner%20just%20announced%20the%20Market,technical%20foundation%20for%20AI%20framework.

¹² RAND Corporation. (2024). Why AI projects fail. <https://www.rand.org/pubs/articles/2024/whyai-projects-fail.html>

¹³ Gartner. (2023). How to stop data quality undermining your business. <https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business>

¹⁴ S&P Global. (2025). 2025 global trends in AI report. <https://www.spglobal.com/market-intelligence/en/news-insights/research/2025-global-trends-in-ai>

¹⁵ DAMA UK Working Group. (2013). The six primary dimensions for data quality assessment. <https://www.dama.org.uk>

¹⁶ Gartner. (2023). How to stop data quality undermining your business. <https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business>

¹⁷ New Relic. (2024). 2024 observability forecast. <https://newrelic.com/resources/report/observability-forecast/2024>

¹⁸ Mordor Intelligence. (2025). Data observability market size, share, and growth analysis & growth report, 2030. <https://www.mordorintelligence.com/industry-reports/data-observability-market>

¹⁹ Gartner. (2024, June 25). Market guide for data observability tools. <https://www.gartner.com/en/documents/5533895>

²⁰ International Data Corporation. (2024). Observability and system resilience research. As cited in Iliia Digital. <https://ilia.digital/en/observability-optimizing-response-time-and-system-resilience/>

²¹ Gartner. (2024, November 19). Gartner forecasts worldwide public cloud end-user spending to total \$723 billion in 2025. <https://www.gartner.com/en/newsroom/press-releases/2024-11-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-total-723-billion-dollars-in-2025>

²³ Precedence Research. (2025). Retrieval augmented generation market size to hit USD 67.42 billion by 2034. <https://www.precedenceresearch.com/retrieval-augmented-generation-market>

²⁴ Gartner. (2024). How intelligent agents in AI can work alone. <https://www.gartner.com/en/articles/intelligent-agent-in-ai>

²⁵ Market.us. (2025). Agentic AI in enterprise IT market size. <https://market.us/report/agentic-ai-in-enterprise-it-market/>

²⁶ Gartner. (2024). Data quality: Why it matters and how to achieve it. <https://www.gartner.com/en/data-analytics/topics/data-quality>

²⁷ Forrester Research. (2023). Data culture and literacy survey. As cited in Grow.com. <https://www.grow.com/blog/the-cost-of-poor-data-quality-and-how-to-avoid-it>

²⁸ IBM. (2024). What are data silos? <https://www.ibm.com/think/topics/data-silos>

²⁹ Dehghani, Z. (2022). Data mesh: Delivering data-driven value at scale. <https://www.oreilly.com/library/view/data-mesh/9781492092384/>

³⁰ Confluent. (2025). 2025 data streaming report: Moving the needle on AI adoption, speed to market, and ROI. <https://report.confluent.io/>

³¹ Dynatrace. (2024). The state of observability 2024: Overcoming complexity through AI-driven analytics and automation strategies. <https://www.dynatrace.com/info/reports/state-of-observability/>

³² Amazon Web Services. (2022). AWS data transfer charges for server and serverless architectures. AWS Partner Network Blog. <https://aws.amazon.com/blogs/apn/aws-data-transfer-charges-for-server-and-serverless-architectures/>

³³ European Union. (2022). Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector (Digital Operational Resilience Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2022/2554/oj>

³⁴ European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

³⁵ McKinsey & Company. (2025). The state of AI in 2025: Agents, innovation, and transformation. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

³⁶ IBM. (2024). What are data silos? IBM Think. <https://www.ibm.com/think/topics/data-silos>

³⁷ Heward Mills. (2025, May 29). How DORA and the EU AI Act are shaping compliance in the financial services sector. <https://www.hewardmills.com/how-dora-and-the-eu-ai-act-are-shaping-compliance-in-the-financial-services-sector/>

³⁸ Gartner. (2025, June 25). Gartner predicts over 40% of agentic AI projects will be canceled by the end of 2027 [Press release]. <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>

³⁹ McKinsey & Company. (2025, October 16). Deploying agentic AI with safety and security: A playbook for technology leaders. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders>

⁴⁰ Gartner. (2024, September 9). Gartner predicts 40% of generative AI solutions will be multimodal by 2027 [Press release]. <https://www.gartner.com/en/newsroom/press-releases/2024-09-09-gartner-predicts-40-percent-of-generative-ai-solutions-will-be-multimodal-by-2027>

Appendix B:

Glossary of Terms

Agentic AI:

AI systems that take autonomous actions with minimal human supervision, making decisions and executing tasks independently based on data inputs.

Data Mesh:

Decentralized data architecture where domain teams own their data as products, promoting self-serve data infrastructure and federated governance.

Data Observability:

Practice of monitoring, tracking, and ensuring the health of data across the entire lifecycle, including quality, freshness, lineage, and schema stability.

Data Product:

An intentionally designed data asset that is discoverable, addressable, understandable, trustworthy, self-describing, interoperable, and secure.

DORA (Digital Operational Resilience Act):

EU regulation requiring financial institutions to demonstrate ICT operational resilience, including data quality and incident response capabilities.

Embedding:

Numerical vector representation of data (text, images, etc.) that captures semantic meaning, enabling similarity search and machine learning operations.

EU AI Act:

European Union regulation establishing requirements for AI systems, including data quality, transparency, and governance obligations for high-risk applications.

FinOps:

Practice of bringing financial accountability to cloud spending through collaboration between engineering, finance, and business teams.

MTTD (Mean Time to Detect):

Average time between when a data quality issue occurs and when it is detected by monitoring systems.

MTTR (Mean Time to Resolve):

Average time between when a data quality issue is detected and when it is fully resolved.

RAG (Retrieval-Augmented Generation):

An AI pattern that grounds large language model responses in retrieved documents, improving accuracy and reducing hallucinations.

Schema Stability:

Measure of how predictable and consistent the data structure remains over time, including column names, data types, and relationships.

Shift-Left:

Practice of moving quality checks earlier in the data lifecycle, catching issues at ingestion rather than after propagation.

Vector Database:

A specialized database designed to store and query high-dimensional vector embeddings, enabling similarity search for AI applications.

Zero-Egress Architecture:

A design pattern where data processing occurs entirely within the data platform, eliminating data movement across network boundaries.

Zombie Pipeline:

An Automated data process that continues to run and consume compute resources after its business purpose has ended.

Appendix C: About DataRadar



DataRadar™ is the only platform that makes your Snowflake data AI-Ready— with data quality monitoring + cost optimization in one Native App.

Trust Your Data.

Control Your Costs.

Power Your AI.



Why DataRadar

Complete data observability requires coverage across five dimensions: Data Reliability, Pipeline Health, Performance Optimization, Usage Intelligence, and Cost Visibility.¹ The market is bifurcated — quality-focused vendors address the first four dimensions (Data Reliability, Pipeline Health, Performance Optimization, and Usage Intelligence), while cost-focused vendors address Cost Visibility alone. DataRadar delivers unified coverage across all five through a single Snowflake Native App.

The DataRadar Difference

As a Snowflake Native App, DataRadar runs entirely inside your Snowflake account—not alongside it.²

	Traditional Vendors	DataRadar
Data Location	Vendor's cloud	Your Snowflake
Data Egress	Required	Zero
Security Review	New vendor assessment	Inherits Snowflake RBAC ³
Deployment	Weeks	In as little as 30 minutes



1. Unified: Data quality monitoring and cost optimization in one Snowflake in one Snowflake Native App, covering The DataRadar™ Observability Framework: Data Reliability, Pipeline Health, Performance Optimization, Usage Intelligence, and Cost Visibility.¹



2. Native: Zero-egress architecture. Your data never leaves your Snowflake environment.²



3. Rapid: Deploy directly from the Snowflake Marketplace in under 30 minutes.



4. Approved: Pay with existing Snowflake committed credits. No new procurement cycle required.

Contact Information



Website:
www.dataradar.io



Email:
Kenneth.Kasee@DataRadar.com



LinkedIn:
[linkedin.com/company/dataradar-io](https://www.linkedin.com/company/dataradar-io)



Snowflake Marketplace:
Search "DataRadar"

Ready to make your data AI-ready? Visit www.dataradar.io to schedule a demo.

References

¹ DataRadar, (2026, May 15). Trust Your Data. Power Your AI. Data Quality + Cost Optimization. <https://www.dataradar.io/>

² Snowflake Inc. (n.d.). About Snowflake Native Apps. <https://docs.snowflake.com/en/developer-guide/native-apps/native-apps-about>

³ Snowflake Inc. (n.d.). Overview of access control. <https://docs.snowflake.com/en/user-guide/security-access-control-overview>

