UNLOCK YOUR DATA INITIATIVES WITH DATAOPS



What AI Demands from DataOps—and How Orchestration Delivers It

DataOps is the practice of applying agile and DevOps principles to data pipeline management—automating, scaling, and operationalizing data workflows across the enterprise. For organizations investing in AI, DataOps has become foundational: AI systems require data that is trusted, validated, and ready for consumption, not just data that moves fast. Closing the gap between pipeline execution and trusted outcomes is now the defining challenge of modern DataOps.

Why do so many data and AI initiatives stall before production?

Most organizations can build a promising proof of concept. The harder problem is getting it to production. According to Forbes, the vast majority of generative AI pilots never make it into production—not because the ideas aren't sound, but because scaling those initiatives requires reliability, integration, and data quality that pilot environments don't demand.

This reflects a broader challenge that extends beyond AI alone. Despite continued investment in data platforms and cloud technologies, organizations still struggle to operationalize data and AI initiatives at scale. The effort remains concentrated on building pipelines and training models. The harder problem—ensuring those systems run consistently in production, with the coordination and data quality required to support business-critical decisions—is where progress stalls.

How DataOps changed the way organizations manage data pipelines

DataOps emerged as a framework to address a fundamental challenge: how to operationalize data initiatives at scale. As data volumes grew and pipelines became more complex, traditional approaches struggled to keep up. DataOps applied agile engineering and DevOps best practices to data management, with a clear goal: turn new insights into production-ready data pipelines that deliver business value.

Adoption accelerated quickly, and DataOps evolved into a recognized market category, Gartner® defines DataOps as the collaborative data management practice focusing on improving communication, continuous integration, automation, observability and operations of data flows between data managers, data consumers, and their teams across the organization. At the center of that definition is orchestration: the coordination, automation, and control that makes complex data workflows manageable.

What AI-ready data actually requires

DataOps helped organizations automate and scale data pipelines—but AI systems have raised the bar. Speed alone is no longer sufficient.

AI-ready data, as we see it in the Gartner report, requires three things:

- **Alignment**—relevance to the specific use case
- **Qualification**—continuous validation for production environments
- **Governance**—policy, compliance, and traceability

This represents a fundamental shift in what DataOps must deliver. DataOps is no longer just about moving data efficiently—it is about ensuring data is ready for intelligent systems to act on.

Why data quality is now an orchestration problem

Many organizations have automated orchestration. Far fewer have implemented data readiness and certification practices. That gap is where AI initiatives break down.

When pipelines prioritize speed over readiness, unvalidated data reaches AI systems and drives business decisions. The result is not just pipeline failure—it is incorrect outcomes at scale. Data quality cannot be treated as a separate, downstream check. It must be embedded within the workflow itself—validating data in context, gating execution based on readiness, preventing downstream impact, and triggering remediation automatically. Without that level of coordination, pipelines may run, but their outputs cannot be trusted.

What modern data pipeline complexity looks like

Modern data pipelines span multiple applications, data sources, and infrastructure technologies that must work together seamlessly. Organizations now operate a mix of analytics pipelines, machine learning pipelines, RAG pipelines, and real-time inference pipelines. While each serves a different purpose, all rely on the same foundation: ingestion, storage, processing, and delivery.

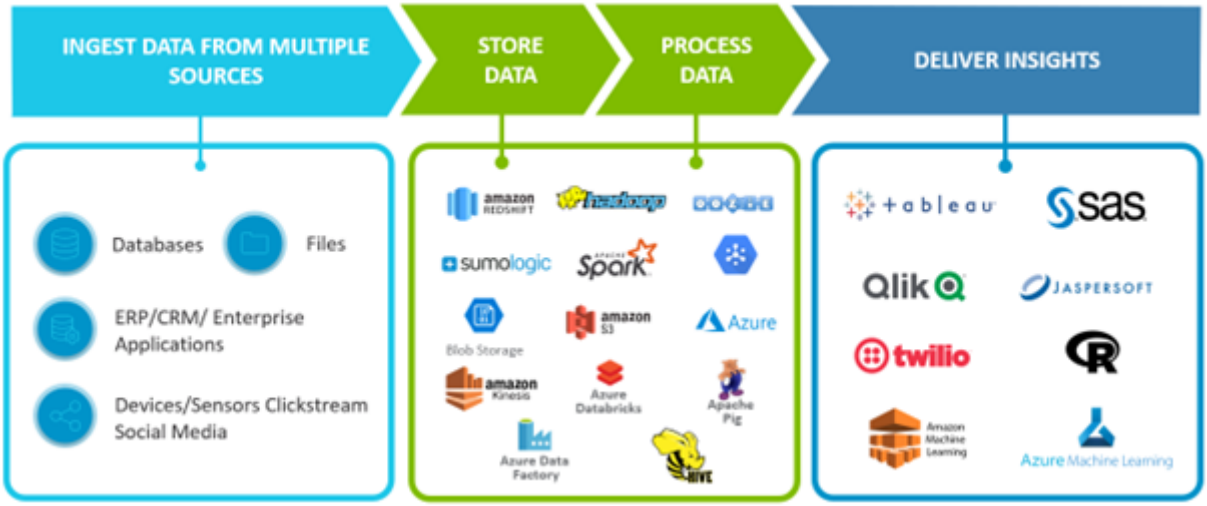


Figure 1. Data projects have four stages with many moving parts across multiple technologies.

Ingestion pulls data from enterprise systems, devices, and digital channels. Storage spans diverse platforms optimized for cost, scale, and access. Processing encompasses batch, real-time, and event-driven workloads—now extended to feature engineering, model training, and inference. Delivery feeds not just dashboards but applications, APIs, and AI models and agents driving automated decisions.

As pipelines become interconnected, data quality is no longer isolated to a single stage. Issues introduced upstream can silently propagate downstream—affecting analytics, model performance, and business decisions. Pipelines may run and SLAs may be met, yet outcomes can still be wrong. In AI-driven systems, this leads to incorrect automated decisions at scale.

This is why orchestration must evolve from a scheduling layer to a control layer—one that determines not just when workflows run, but whether they should run at all, based on data readiness and quality.

Operationalizing DataOps for the AI era with Control-M

DataOps provides the foundation for managing data pipelines. But success in AI-driven environments depends on the ability to operationalize data, AI, and application workflows as a unified system—reliably and at scale. The gaps identified above—data readiness, execution gating, event-driven coordination, and cross-platform visibility—require a control layer that operates above individual pipelines. Control-M and Control-M SaaS provide that layer.

By abstracting the complexity of modern environments, Control-M orchestrates workflows across data platforms, AI systems, and business applications—providing end-to-end visibility, predictive SLAs, and coordinated execution across hybrid and multi-cloud environments.



Figure 2. Control-M is a layer of abstraction to simplify complex data pipelines.

Orchestrating across the modern data and AI ecosystem

Modern workflows span a rapidly expanding ecosystem of technologies. Control-M provides deep, out-of-the-box integrations across data platforms such as Snowflake, Databricks, and BigQuery; AI platforms including SageMaker, Bedrock, and Vertex AI; streaming systems like Kafka, SQS, and RabbitMQ; and enterprise applications across ERP, CRM, and modern SaaS platforms. This breadth allows organizations to orchestrate end-to-end workflows across the entire data and AI stack—without relying on custom scripting or fragmented tools.

Enabling event-driven and real-time workflows

AI-driven systems increasingly operate in real time—triggered by events, not schedules. Control-M supports event-driven orchestration, enabling workflows to be initiated based on data changes, application events, and business triggers. This enables more responsive execution for inference pipelines, agent-driven decisions, and time-sensitive processes.

Embedding data readiness with Control-M Data Assurance

In AI-driven environments, execution without validation introduces risk. Control-M Data Assurance embeds continuous data validation directly into workflows, ensuring data is validated before, during, and after execution. Workflows can be gated based on data readiness, preventing bad data from triggering downstream processes and ensuring that AI models and agents operate only on trusted inputs. This shifts DataOps from pipeline execution to trusted execution.

SLA-driven, outcome-aware orchestration

Control-M connects workflow execution to business outcomes—tracking SLAs, predicting delays, and surfacing impact—so workflows align to business-critical timelines.

Automation, remediation, and visibility at scale

Control-M automates error detection, notifications, remediation, and cross-platform coordination,

while providing end-to-end visibility across workflows for operations teams, developers, and business users.

The need for data continues to grow—but value is only realized when that data is operationalized effectively. By combining DataOps practices with a control layer like Control-M, organizations can orchestrate data, AI, and application workflows end to end, integrate across evolving ecosystems, respond to events in real time, ensure data readiness before execution, and deliver reliable, SLA-driven outcomes at scale.

Success in the AI era is not defined by how fast pipelines run—but by how reliably they produce trusted, actionable outcomes.

Frequently asked questions about DataOps and AI orchestration

What is the difference between DataOps and traditional data management?

Traditional data management focuses on storing and accessing data. DataOps applies agile engineering and DevOps principles to the full lifecycle of data pipelines—automating delivery, improving collaboration between teams, and enabling faster, more reliable production deployment of data workflows.

Why do generative AI pilots often fail to reach production?

Most generative AI pilots fail to reach production not because the underlying technology is flawed, but because scaling AI requires data that is continuously validated, reliably delivered, and governed for compliance—conditions that pilot environments rarely test for. Without production-grade DataOps practices in place, AI systems encounter data quality issues that cause failures or incorrect outcomes at scale.

What is AI-ready data?

AI-ready data meets three criteria that we see in the Gartner Report: alignment (relevance to the specific use case), qualification (continuous validation for production environments), and governance (policy, compliance, and traceability). Data that moves quickly but fails these criteria can produce incorrect outcomes when consumed by AI systems.

What is the role of orchestration in DataOps?

Orchestration is the coordination layer that manages when, how, and whether data workflows execute. In modern DataOps, orchestration has expanded beyond scheduling to include data readiness gating, event-driven triggering, SLA tracking, and automated remediation—ensuring that pipelines not only run, but produce trusted outcomes.

How does Control-M support DataOps for AI?

Control-M acts as a control layer for data, AI, and application workflows—orchestrating execution across platforms such as Snowflake, Databricks, SageMaker, and Kafka. Control-M Data Assurance embeds data validation directly into workflows, gating execution based on readiness so that AI models and agents operate only on trusted inputs.

What is the difference between pipeline orchestration and a control layer?

Pipeline orchestration coordinates task execution within a workflow. A control layer operates at a higher level—determining whether workflows should run based on data quality and readiness, coordinating across multiple systems and pipeline types, and connecting execution to business outcomes through SLA tracking and predictive monitoring.

Kindly add the required attributions and disclaimers –

Gartner, Market Guide for DataOps Tools, By Michael Simone, Sharat Menon, Robert Thanaraj, 24 October 2025.

GARTNER is a trademark of Gartner, Inc. and/or its affiliates.

Gartner does not endorse any vendor, product or service depicted in its research publications and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.