Market Guide for DataOps Tools

24 October 2025 - ID G00836742 - 37 min read

By: Michael Simone, Sharat Menon, Robert Thanaraj

Initiatives:Data Management Solutions

Data and analytics leaders face growing pressure to optimize data delivery across increasingly fragmented environments, driven by the urgent need for AI-ready data. This guide helps you navigate the DataOps tools market and select solutions that streamline orchestration, observability, automation, and testing.

Overview

Key Findings

- Technology silos impede organizations from delivering trusted, secure, high-quality, and consumable data on time. DataOps tools reduce friction points through data pipeline orchestration, automation, and observability in complex environments.
- DataOps is gaining momentum as a foundational strategic enabler of streamlined data delivery. Capabilities such as scalable orchestration, continuous data quality, and contextual governance enforcement are now essential.
- AI-assisted and agentic-enabled DataOps tools are emerging to automate manual tasks across data pipeline development and orchestration.
- Feature convergence across data management solutions creates buyer confusion. Overlapping capabilities among DataOps, data observability, and data integration tools often lead to redundant functionalities and unclear differentiation.
- The rise of data management platforms (DMPs) adds to decision complexity by blending overlapping unified capabilities like pipeline orchestration, observability, deployment automation, and test automation.

Recommendations

- Drive velocity across workflow integration, task automation, and pipeline reliability by evaluating AI-assisted DataOps tools and supporting protocols like Model Context Protocol (MCP) and agent-to-agent (A2A) for intelligent agentic coordination.
- Streamline tool selection by mapping redundancies across DataOps, observability, and integration platforms, then prioritize modular, interoperable solutions that can flexibly integrate and interoperate with each other.
- As buyers consolidate around integrated platforms, evaluate whether embedded DataOps features are “good enough.” Use stand-alone tools strategically in multidomain, multicloud, or specialized scenarios where platform-native capabilities fall short.
- D&A leaders must evaluate those DataOps vendors that deliver unified data operations and observability capabilities across pipelines, infrastructure, orchestration, financials, and lineage.
- Identify the embedded DataOps capabilities of your strategic long-term data management platforms (DMP) and assess if those are adequate for your current data delivery use cases.Only augment/complement with a stand-alone DataOps tool if the DMP does not provide DataOps functionality that is good enough for your needs.
- Identify embedded DataOps capabilities in your strategic data management platforms (DMPs) and assess their fit for current delivery needs. Augment or supplement with stand-alone tools only if platform-native functions fall short.

Market Definition

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. DataOps tools connect and orchestrate data pipelines across heterogeneous systems. Data and analytics leaders are the buyers in this emerging market. The primary audience for DataOps tools is “data manager” personas like, data engineers, data integration developers, operations/incident analysts, database administrators and data architects. The secondary audience is “data consumer” personas like business analysts, business intelligence developers, data scientists and citizen roles (departmental users who are domain experts, but less technical).

Data and analytics leaders seeking to deliver high-quality, trusted data products to their businesses should invest in a DataOps tool to streamline their data delivery operations and reduce the manual effort required to manage complex data pipelines. This class of tools provides greater control, transparency, and agility over the full life cycle of data pipeline management. DataOps tools specifically focus on the end-to-end flow of data.

Common use cases include:

- **Data pipeline orchestration and monitoring:** Improve the speed of delivery of trusted data products to business consumers by optimally orchestrating and monitoring the execution of end-to-end data flows to better coordinate and monitor data pipeline workloads across heterogeneous data stacks. This enables data teams to meet data delivery SLA commitments.
- **Data pipeline observability:** By continuously interrogating data structures, pipelines, and their contents, DataOps tools provide early-warning alerts that surface data exceptions and pattern deviations throughout the data pipeline life cycle. This awareness gives data teams the ability to remediate data and data processes earlier, thereby preventing poor-quality data from being delivered to downstream businesses that could negatively impact their processes.
- **Environment management:** In order to execute repeatable testing patterns across environments (like development, testing and production) DataOps tools can reliably and repeatedly instantiate templated infrastructure-as-code build specs across data pipeline development life cycle stages. This ensures that the ramifications of various environment runtime conditions are consistently measured and applied to meet production business delivery requirements.

Must-Have Capabilities

The must-have capabilities for vendors in this market include:

- **Data pipeline orchestration:** Automates, coordinates and monitors existing and new data pipelines from a single integrated platform simplifying the management of end-to-end data flow complexity through a common control plane. It manages, integrates, schedules and automates tasks across diverse sets of data pipelines. Includes connector management, workflow impact analysis, and data pipeline audit logs.

- **Data pipeline observability:** Optimizes pipelines, monitors and detects data quality issues, collects data usage patterns, traces data flow lineage, and spots data anomalies using native and/or integrated live and historical metadata that is harvested from orchestrated pipelines and related data assets. This includes monitoring, logging, business-rule detection and metadata capture.

Common Features

The standard capabilities for vendors in this market include:

- **Data pipeline deployment automation:** Automates the version-control processes, DevOps integration, release of pipelines, and change management approvals of data-related components across release cycles via a continuous integration/continuous deployment (CI/CD) methodology.
- **Data pipeline test automation:** Practices executions of pipeline code "dry runs," business rules validation, test scripts management, regression testing packs and test data management.
- **Environment management:** Minimize the repeated manual effort creating, maintaining, and optimizing the deployment of pipelines across environments (development, testing (including user acceptance testing), staging and production, for example). This uses an infrastructure-as-code model to execute repeatable testing patterns including resource provisioning, environment templating and secure credentials. DataOps can reliably and repeatedly instantiate build specs across all data pipeline life cycle stages ensuring that environment runtime conditions are consistently applied.

Market Description

Organizations today operate many data and analytics tools to support increasingly diverse business goals. However, these tools often function in silos that are managed by vertically aligned expert teams using disjointed development environments and asynchronous life cycles. This fragmentation introduces significant data delivery challenges, including slow time-to-market, inconsistent environments, release bottlenecks, governance gaps, and complex dependency handoffs.

In the last year, embedded DataOps tool capabilities in other platforms have started to shift buyer preferences and intensified competitive pressure on stand-alone vendors. As orchestration, observability, and automation become baseline expectations, stand-alone DataOps solutions risk being viewed as redundant. Buyers are increasingly leaning toward integrated platforms, though stand-alone tools still hold strategic value in specific contexts.

DataOps tools orchestrate critical pipeline stages such as ingestion, transformation, testing, and deployment while reducing manual effort and operational friction. By doing so, they enable scalable, reliable, and efficient data delivery that meets the evolving needs of data consumers across the organization. Figure 1 highlights the scope of DataOps tools across five core capability sets:

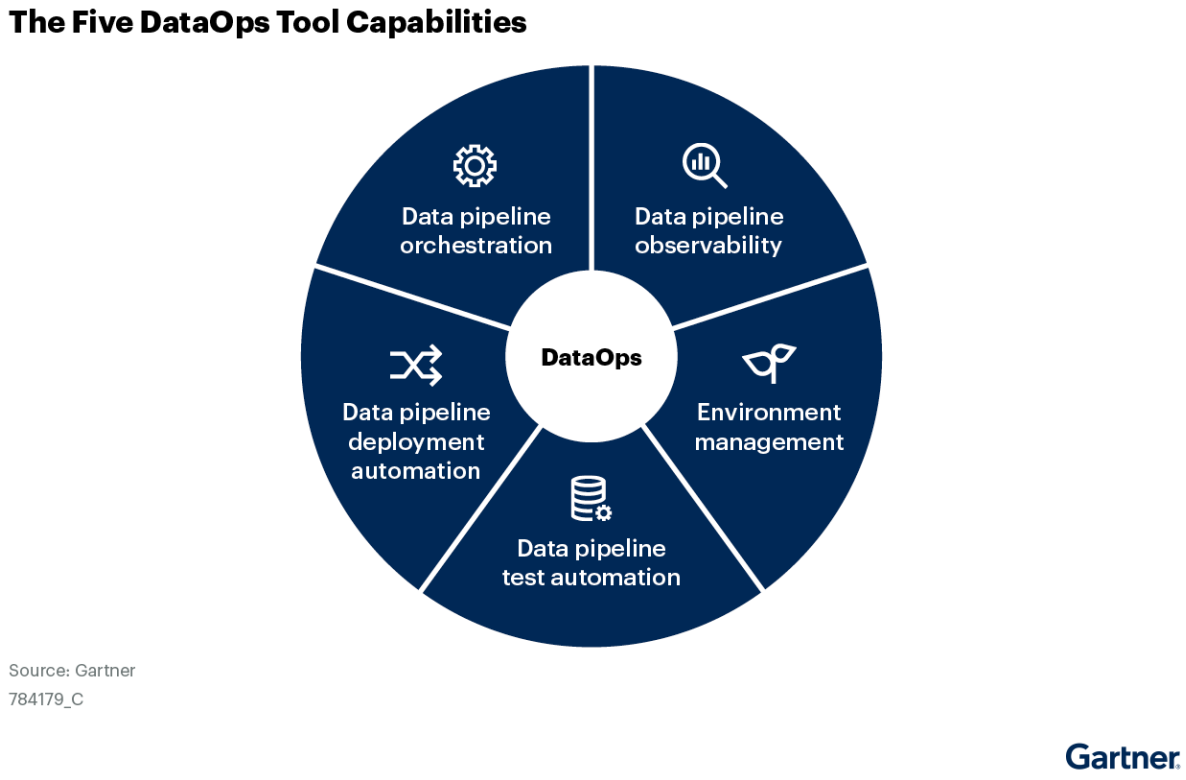
- Data pipeline orchestration
- Data pipeline observability
- Data pipeline test automation
- Data pipeline deployment automation
- Environment management

DataOps tools allow data engineers to automate the pipeline tasks within these steps to better support data consumers in meeting their business data goals. Much like DevOps transformed software engineering, DataOps fosters cross-functional collaboration among data engineers, data managers, and business stakeholders by providing a unified experience across the data life cycle.

DataOps tools address inefficiencies and misalignments between data management and data consumption by streamlining data delivery, aligning cross-functional data team personas, and operationalizing key data artifacts, including processes, pipelines, data assets, and platforms. This integrated approach accelerates the delivery of business value by enabling faster, more reliable, and outcome-driven data operations.

— *Gartner*

Figure 1: The Five DataOps Tool Capabilities



Gartner

Market Direction

In the 10 years since DataOps first emerged, vendors have continued to grow with updated and enhanced product lines of varied scope focused on different combinations of data pipeline orchestration and observability, along with other capabilities such as data pipeline testing, automation, quality validations, and environment management.

Despite extending varied capabilities, DataOps vendors face barriers to entry from incumbent tools already deployed in organizations and the growing availability of DataOps features in adjacent markets. They struggle to overcome challenges from decision makers who do not recognize the additional benefits beyond their organization's incumbent tools and data market convergence. This accelerates the commoditization of DataOps capabilities, as converged platforms subsume more core functions, making it harder for DataOps vendors to differentiate and justify premium value.

These buyer decision considerations are reshaping the competitive landscape. Traditional data management vendors increasingly embed DataOps capabilities into their platforms, spanning data integration, quality, metadata, and master data management, creating broader, more unified offerings. This is also accelerating as orchestration, automation, and quality features become standard across platforms, driven by the rising need to support AI-ready data. For example, platforms initially focused on data pipeline orchestration are now expanding into adjacent capabilities such as data observability and environment management.

This growing overlap between DataOps and adjacent markets has led to increasingly complex platform buying decisions. The following section illustrates how these overlaps unfold, highlighting their implications for DataOps capabilities, tooling strategy, platform differentiation, and buyer decision making.

- Data observability tools increasingly overlap with DataOps, particularly data pipeline monitoring. D&A leaders should expect continued feature intersections driven by the ongoing needs for broader observability insights across data environments (see Market Guide for Data Observability Tools).

- Emerging data management platforms (DMPs) integrate DataOps into a unified solution. This consolidation appeals to buyers seeking simplified vendor relationships and unified data capabilities on a single platform, which is intensifying competitive pressure on stand-alone DataOps vendors (see Market Guide for Data Management Platforms).

- GenAI, natural language processing (NLP), ML, agentic AI, and autonomous orchestration are reshaping the DataOps market by simplifying and automating data workflows. This is shifting buyer expectations toward intelligent solutions that reduce manual effort and deliver faster realization of business goals.

- Data integration vendors have increasingly embedded DataOps capabilities into their solutions, entering the DataOps tool market to meet buyer demand for unified solutions that streamline data pipeline management and reduce tooling fragmentation (see Magic Quadrant for Data Integration Tools).

The DataOps tools market grew by 21% in 2024 to reach a size of $424M (see Market Share Analysis: Data Management Software (Excluding DBMS), Worldwide, 2024), well above the overall data management software growth of 9%. Gartner expects the market to continue growing as a stand-alone market for the next one to two years, driven by the need for cross-platform data orchestration and observability.

Increasingly, DataOps vendors will face strong pressure to differentiate themselves against emerging DMPs with embedded DataOps capabilities. As this trend accelerates, stand-alone DataOps tools will likely be assimilated into broader data management platforms as embedded capabilities, reducing the need for separate purchases. The only compelling reasons for users to continue buying stand-alone DataOps tooling will be:

- Multicloud hybrid scenarios, where embedded DataOps capabilities within cloud platforms do not suffice
- Specialist domain, vertical, or use-case scenarios — such as IoT — where embedded capabilities within DMPs may fall short
- Specialist DataOps tools that support the full end-to-end functionality of DataOps, which are often fragmented across three to four separate tools within CDMPs

In the near term, D&A leaders should evaluate their investment in DataOps tools by aligning data engineering needs with the complexity of existing data management environments. While DMPs, observability tools, and integration platforms offer overlapping features, they often lack the orchestration, automation, and depth engineers need to manage pipelines efficiently across hybrid architectures and CI/CD workflows.

The DataOps market has evolved into a strategic data management foundational capability, driven by the increased need for optimized data pipelines to support AI initiatives and growing analytic use cases.

Market Analysis

Gartner's 2024 Evolution of Data Management Survey reports that 52% of respondents indicate that they have implemented DataOps tools. Reinforcing this maturity, the 2025 Gartner Hype Cycle for Data Management shows the DataOps tools category progressing from adolescent to early mainstream, reflecting sustained demand and broader adoption.

DataOps tools support diverse processing patterns and deployment models, including operational and analytical workloads, AI-ready data preparation, batch and real-time flows, and environments spanning on-premises and cloud. Some vendors offer modular SKUs, with distinct capabilities priced separately, which can impact the total cost of ownership and deployment flexibility.

DataOps vendors generally fall into three categories based on their core focus:

- **Generalists:** These tools optimize end-to-end data pipeline workflows by combining the five core DataOps capabilities to varying extents. Choose generalists for broad, end-to-end data flow management DataOps needs.
- **Specialists:** These tools target specific pipeline gaps, such as infrastructure cost optimization, automated data validation, or standardized testing. Choose specialists to address focused needs within existing pipelines.
- **Orchestrators:** These tools prioritize data orchestration and job scheduling, often adding features like CI/CD and observability. Choose orchestrators for job/task scheduling and coordinated management of complex, multipipeline workloads.

Despite market fragmentation, demand among data and analytics leaders is rising to manage increasing operational complexity. Gartner inquiry data shows consistent interest in DataOps tools and practices, with over 1,000 client interactions from Q1 2024 through Q3 2025. See Note 1 for TAM estimates of the DataOps technology market.

As buyer priorities shift toward agility and intelligent automation, DataOps is gaining momentum as a foundational strategic enabler of streamlined data delivery (see Streamline Data Delivery Using 5 Essential DataOps Practices). Capabilities such as scalable orchestration, continuous data quality, and contextual governance enforcement are now essential.

Within the next 12 to 24 months, AI-assisted and agentic capabilities will become baseline expectations for DataOps platforms, enabling automated pipeline generation, governance operations, and self-healing optimization. While these features will define short-term competitiveness, sustainable differentiation will depend on vendors delivering deep semantic interoperability, autonomous policy enforcement across federated environments, and intelligent runtime frameworks capable of continuous self-optimization in response to dynamic execution environments and technology shifts.

Buyers are quickly prioritizing solutions that drive velocity of scalable, governed, and AI-ready pipelines, not just automation (see Use These 5 DataOps Practices to Drive Collaboration and Governance).

Buyers are also beginning to consolidate around data management platforms (DMPs) to reduce complexity. They increasingly expect integrated DataOps capabilities directly into these platforms, forcing DataOps vendors to choose between further specialization and/or expansion into broader platform ecosystems, strategic alignment with broader platforms through partnerships, integrations, or native embedding.

Market confusion stems from the varied placement of DataOps functions across the software stack. Data integration tools increasingly offer overlapping pipeline orchestration, observability, optimization, and version control (see Critical Capabilities for Data Integration Tools). Service orchestration platforms also enter the DataOps space by adding observability features (see Critical Capabilities for Service Orchestration and Automation Platforms).

These trends are driving vendor repositioning. To remain competitive, DataOps vendors must deliver unified data operations and observability capabilities across pipelines, infrastructure, orchestration, financials, and lineage. Those that fail to do so risk falling behind competitors that offer AI-enabled data operations as buyers prioritize automation and intelligence.

Representative Vendors

The vendors listed in this Market Guide do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings.

Market Introduction

The vendors named in this guide are listed in alphabetical order. They were selected to provide a representative range of DataOps tools and do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings. Table 1 below summarizes these vendor offerings. Gartner received details on geographic presence, industry coverage, and representative customers directly from these vendors. Industry coverage is not listed in any particular order.

Table 1: Representative Vendors in the DataOps Market

(Enlarged table in Appendix)

Vendor	Product Names	DataOps Category	Geographic Presence	Sample of Representative Industry Coverage
Acceldata	Data Observability Cloud	Specialist (data pipeline observability)	All regions	Financial, telco, consumer packaged goods (CPG), technology
Ascend.io	Data Automation Cloud	Generalist	North America, Europe, APAC	Media, technology, healthcare, retail, financial services
Astronomer	Astro, Astro Observe	Orchestrator	All regions	Financial services, insurance, manufacturing, healthcare, retail
Amazon Web Services (AWS)	Amazon Managed Workflows for Apache Airflow (MWAA)	Orchestrator	All regions	Advertising, automotive, manufacturing, utilities, financial services
BMC Software	Control-M, BMC Helix Control-M	Orchestrator	Americas, Europe, Asia/Pacific	Financial services, telco, retail, logistics, government
Composable Analytics	Composable DataOps Platform	Generalist	North America	Financial services, pharma, telco, utilities, government
Dagster Labs	Dagster+	Orchestrator	North America, Europe, APAC	Financial services, pharma, energy, retail, utilities
Datagaps	DataOps Suite, ETL Validator, BI Validator, DQ Monitor, Test Data Mgr.	Specialist (data pipeline test automation)	North America, Europe, Australia, India	Financial services, healthcare, education, energy, technology, manufacturing, telco
DataKitchen	DataOps Observability, DataOps TestGen, DataOps Automation	Generalist	Americas, Europe	Utilities, financial services, life sciences, manufacturing, healthcare
DataOps.live	DataOps.live	Generalist	U.S., Europe	Communications, technology, media, healthcare, life sciences
Google	Cloud Composer	Orchestrator	All regions	Healthcare, retail, financial services, media, manufacturing
IBM	watsonx.data intelligence	Generalist	All regions	Financial services, government, insurance, industrial, healthcare
iceDQ	iceDQ	Specialist (Data Pipeline test automation)	North America, Europe, APAC	Utilities, Financial Services, government, insurance, manufacturing
lakeFS	lakeFS	Specialist (Environment Management, Test automation)	North America, Europe	Financial services, healthcare, manufacturing, utilities
Microsoft	Azure Workflow Orchestration Manager	Orchestrator	All regions	Not available
Prefect	Prefect Cloud	Orchestrator	Americas, EMEA, APAC	Financial services, insurance, energy, education
Stonebranch	Universal Automation Center (UAC), Universal Data Mover Gateway (UDMG)	Orchestrator	Americas, EMEA, APAC	Financial services, insurance, Pharma, retail, manufacturing
Unravel Data	Unravel for Databricks, Unravel for Snowflake, Unravel for BigQuery, Unravel for Amazon EMR, Unravel for Coudera	Specialist (data pipeline observability)	North America, Western Europe, APAC	Financial services, healthcare, life sciences, retail, telco

Source: Gartner

The next section gives D&A leaders an overview of the vendors listed in Table 1.

Vendor Profiles

Acceldata

Acceldata was founded in 2018 and is headquartered in Campbell, California, U.S. Its DataOps product is Acceldata Data Observability Cloud. Acceldata offers two SKUs: Data Quality & Reliability for monitoring data and pipelines, and Cost Optimization for tracking infrastructure and associated expenses.

Acceldata is building its platform around three core pillars — discovering data, enforcing data policies, and unearthing data insights. Acceldata has strategic partnerships with Reltio to enhance MDM support for joint clients, with Fiddler for AI observability, and with global system integrators like Accenture, Cognizant, and Wipro to improve its geographic reach.

In 2025, Acceldata launched Agentic Data Management, enabling natural language interactions and task-specific agents (e.g., metadata, profiling, tagging). These capabilities leverage Amazon Bedrock for GenAI. Integration with Apache Iceberg, Databricks Unity Catalog, and OpenLineage ensures openness and interoperability in client data architectures.

Acceldata monitors pipelines built in multiple languages (e.g., Python, Java) and tools (e.g., Airflow, Azure Data Factory). For orchestration, deployment, and test automation, it partners with Ascend.io and Promethium, while supporting observation of externally built pipelines.

Acceldata's observability capabilities, such as identifying inefficient clusters, mapping cost centers, and tagging assets, enable organizations to strengthen cloud financial management.

Representative customers: Dun & Bradstreet, Nestlé, PhonePe, PubMatic, The Hershey Company

Amazon Web Services (AWS)

Amazon Web Services (AWS) was founded in 2006 and is headquartered in Seattle, U.S. It offers Amazon Managed Workflow for Apache Airflow (MWAA) as a key DataOps component, alongside other AWS services offering DataOps capabilities, including AWS Glue, AWS Step Functions, and Amazon CloudWatch.

Amazon MWAA is a managed orchestration service for Apache Airflow that allows data engineers to execute data processing workflows in the cloud. With Amazon MWAA, developers can offload the operational burden of managing the underlying Airflow infrastructure. It can simplify capacity planning by automatically resizing the compute resource associated with the Airflow worker nodes. It integrates with AWS security services to enable secure access. It provides easy access to the Airflow user interface, which allows developers to observe their workflow executions to determine if there is a problem before performing expensive data transformations or machine learning (ML) training.

MWAA interacts with other AWS services like Amazon S3, Amazon CloudWatch, and AWS CloudFormation.

In 2025, AWS extended the use of Amazon MWAA to Amazon SageMaker for deploying and scaling Airflow DAGs while reducing the infrastructure operational burdens. It has added automated scaling and built-in fault tolerance for MWAA in Amazon SageMaker.

Beyond workflows, AWS's DataOps capabilities provide monitoring, git integration, support for developer tools, and the ability to move between environments.

Representative customers: Choice Hotels, The Walt Disney Company, Itau, Merck, Yahoo Advertising

Ascend.io

Ascend.io was founded in 2015 and is headquartered in Palo Alto, California, U.S. Its DataOps product is the Agentic Data Engineering Platform.

Ascend's platform provides pipeline ingestion, transformation, orchestration, and observability capabilities, including DAG traversal, multi-DAG links with workflow dependency identification, pipeline checkpointing with automated pause, resume, and rollback. Pipelines are built and tested in isolated developer workspaces within the product (in Python and SQL) and deployed to production, where the code is orchestrated and pushed down to cloud warehouses such as Snowflake, Databricks, and BigQuery.

Pipelines built within Ascend and externally can both be orchestrated. Pipelines are versioned using Git for code history. All code is fingerprinted and correlated to data partitions, enabling DataAware Automation that intelligently optimizes data workloads and only reprocesses data as needed based on logic or data changes.

Ascend provides pipeline observability with profiling, lineage, and alerting via a unified metadata layer. It includes pipeline test automation, components, code, columns, and live data quality checks. Runtime parameters enable environment configuration for development, testing, and production

In 2025, Ascend released the AI agent Otto and an agentic data engineering architecture and event-driven framework. Otto works alongside Custom Agents that teams can create for domain-specific workflows and supports code generation, error handling, incident triage, documentation, and pipeline optimization to meet DataOps SLAs.

Representative customers: Biome Analytics, Harry's, Mattel, News Corp., United Healthcare

Astronomer

Astronomer was founded in 2018 and is headquartered in New York, U.S. Astro is available as downloadable software or a cloud service, deployable across on-premises, multicloud, and SaaS environments.

Astronomer Astro delivers a fully managed Apache Airflow service with day zero support for the latest version of Airflow, DAG versioning, GitHub integration, remote execution, in-place upgrades, and deployment rollbacks. It orchestrates external pipelines (Databricks, Snowflake, Fivetran, dbt, Kafka, OpenAI) and is among the first to offer managed Airflow 3.0.

Astro automates data pipeline orchestration, including workflow dependencies, scheduling, monitoring, impact analysis, and alerting. GenAI and LLM workflows are supported for embedding refreshes, retrieval-augmented generation (RAG), AI inference execution, model training/evaluation, and model drift detection. The new browser-based IDE enables users to write, test, and deploy Airflow DAGs without local setup. It includes AI-assisted code generation that produces Airflow-compliant DAGs tailored to project history, Airflow version, and environment.

Astro Observe is a centralized observability layer offering cross-environment pipeline visibility, enabling lineage tracking, anomaly detection, SLA monitoring, and root cause analysis, with predictive analytics to forecast SLA breaches before failures occur.

In 2025, Astronomer introduced AI-assisted log summarization and incident reporting, which automatically pinpoints failure points and suggests fixes for faster debugging. It also leverages large language models for sentiment analysis, similarity search, and text summarization.

Representative customers: Adobe, Condé Nast, FanDuel, Marriott, T. Rowe Price

BMC Software

BMC Software was founded in 1980 and is headquartered in Houston, U.S. It offers Control-M and BMC Helix Control-M as its DataOps products.

BMC introduced Unified View, which enables orchestration across both SaaS and on-prem environments from a single interface, simplifying hybrid pipeline management. Developers can track jobs with dependencies across pipelines through Control-M's monitoring domain. It can visually represent pipelines as a business service and enable tracking SLAs with built-in predictive SLA monitoring and alerting.

Control-M's data pipeline orchestration capabilities include version control, workflow automation, dependency mapping, scheduling, monitoring, impact analysis, and real-time alerting. Observability includes lineage, data profiling, and threshold management.

Automated data validation is enabled directly in AI workflows with Data Assurance. It applies data quality validation metrics to data before it enters pipelines, ensuring reliable, high-quality data and triggering alerts if checks fail.

In 2025, BMC added Jett, a GenAI assistant in Control-M SaaS, enabling natural language workflow interaction for issue resolution, compliance, and performance optimization. BMC also introduced agentic AI via HelixGPT agents, including Insight Finder for AIOps insights and Change Risk Advisor for risk assessment and mitigation.

Control-M is available as self-hosted on single/multiple clouds, hybrid, as well as BMC Helix Control-M, which is a SaaS offering.

Representative customers: CARFAX, Domino's Pizza Group, Marisa S.A., Navistar, Railinc

Composable Analytics

Composable Analytics was founded in 2014 as a spinoff from MIT's Informatics and Decision Support Group. It is headquartered in Cambridge, Massachusetts, U.S. Its DataOps product is named Composable DataOps Platform.

Composable DataOps Platform supports data pipeline orchestration and observability DataOps capabilities on a self-contained platform. Composable enables self-service data pipeline and workflow development via a low-code web-based integrated development environment (IDE).

Composable has introduced agentic AI capabilities that enable autonomous orchestration of data workflows. These capabilities allow intelligent agents to manage tasks like extraction, transformation, and model training. These agents operate within a composable architecture that supports multiagent collaboration, semantic enrichment, and real-time decision making.

The Test Suites framework supports debugging, workflow optimization, and real-time data capabilities additions for validating pipeline integrity, as well as Network View for visualizing relationships among data flows and resources. The platform now supports unit testing for data flows, video/audio rendering in pipeline runs, and new modules such as Kafka message readers and speech-to-text processors. Composable offers orchestration for both native and external components, with CI/CD support and infrastructure-as-code for managing compute and storage.

Its ML-driven orchestration supports MLOps and ModelOps. Observability features include workflow monitoring, data lineage, resource usage, and impact analysis, with historical metrics for trend analysis and real-time dashboards for tracking pipeline execution.

Representative customers: Manulife, John Hancock, USAA Real Estate, Takeda

Dagster Labs

Dagster Labs was founded in 2018 and is headquartered in San Francisco, U.S. It offers two DataOps products: open-source Dagster and commercial Dagster+.

Dagster is a code-first orchestration platform built specifically for data engineers. Dagster data pipeline orchestration caters to data pipeline orchestration tasks, including runs, re-executions, retries, partitioning, backfills, and scheduling. It also manages the infrastructure for the runs. All tasks are heavily code-centric (use Python and YAML), from managing pipeline logic to resources and scheduling. Dagster Components are configurable and reusable pipeline building blocks, which enable developers to declare data assets, manage resources, and schedule pipelines.

Dagster+ enables the creation of custom data quality checks and automatically captures lineage with operational metadata like partition health and check status. Data consumers can search for products with rich filters, while asset pages show descriptions, lineage, tags, and owners. It uses anomaly detection to flag overdue updates, block downstream errors, and send alerts. Cost-impact analysis and budget alerts are also supported.

Dagster+ doesn't auto-manage test data, but users can configure it via code. It offers trend reporting for cost and platform optimization and is available in serverless or hybrid modes, supporting on-prem, cloud, and multicloud environments.

In 2025, Dagster released its MCP server, enabling AI assistants like Cursor to integrate with Dagster projects and LLM routing within Dagster pipelines.

Representative customers: Clippd, easyJet Holidays, EvolutionIQ, Group1001, Vanta

Datagaps

Datagaps was founded in 2013 and is headquartered in Herndon, Virginia, U.S. Its DataOps product is called Datagaps DataOps Suite. It is a unified platform that includes an ETL Validator, BI Validator, Data Quality Monitor, and Test Data Manager.

Datagaps offers a suite of data pipeline orchestration and data test automation tools, including orchestration and observability. Recent releases extend orchestration across environments with dedicated engine assignment per container with added Webhook integrations (e.g., Jenkins/Azure DevOps/Slack) and email/Teams notifications. Observability and validations span the end-to-end pipeline life cycle, with profiling, lineage, and versioned workflow tracking for traceability and change management

Newly introduced AI agents generate test cases and rules using ETL mapping documents or code. The Data Rule Wizard accelerates rule creation, and Test Data Manager uses ML to synthesize realistic, privacy-safe test data.

ETL Validator utilizes Azure OpenAI to auto-generate SQL and enrich table/column metadata, while scaling on distributed engines (e.g., Spark), keeping data and logs within the local environment.

Data quality checks can be applied from ingestion through delivery. BI Validator automates regression and filter-level testing, compares dashboards across environments, validates reports against database results, and runs performance/stress tests with AI-generated difference summaries.

Representative customers: Broadcom, Capital One, Sony, UnitedHealthcare, University of Virginia

DataKitchen

DataKitchen was founded in 2014 and is headquartered in Lexington, Massachusetts, U.S. Its DataOps products are DataOps Observability, DataOps TestGen, and DataOps Automation. It has recently open-sourced Observability and TestGen for individuals (data engineers), alongside the paid edition for teams and enterprises.

A user-based pricing model makes it easy for smaller data engineering teams to get started without significant capital investment. It goes beyond selling software by providing DataOps training courses and pragmatic guidance to data professionals through books such as the DataOps Cookbook and the DataOps Manifesto.

DataKitchen TestGen can profile a new dataset and map the appropriate data quality rules (from a library of over 30 rule types) based on the profiling. The vendor then guides engineers on best practices to efficiently validate the datasets that couldn't be auto-validated in the platform itself. Therefore, organizations looking for a lean and agile vendor that can support their data engineers via a balanced software and services-led approach should evaluate DataKitchen.

DataKitchen's DataOps Observability tracks the entire data journey from source to consumption. It offers governed, on-demand workspaces (kitchens) for collaborative environment management, supports Git-based branching, and enables CI/CD through integrations like Jenkins and Ansible.

Representative customers: Bristol Myers Squibb, Catholic Release Services, CNH, Eisai, EQT

DataOps.live

DataOps.live was founded in 2020 and is headquartered in London, England. Its DataOps product is also named DataOps.live.

DataOps.live provides a DataOps automation platform that enables organizations to continuously deliver trusted, governed, and AI-ready data across Snowflake and other cloud ecosystems. The DataOps.live Momentum platform automates the full life cycle of data pipeline development and operations — combining automated CI/CD, continuous observability, governance enforcement, AI-Ready Scoring, and Data Product Delivery within a single, unified environment.

DataOps.live is built natively on Snowflake and extends its capabilities with two distinct families of Snowflake Native Apps. The first, Dynamic Delivery, introduces AI-driven continuous integration and continuous delivery (CI/CD), enabling automated pipeline creation, testing, and deployment across Snowflake environments. The second, Dynamic Transformation, provides orchestration and automation of data transformations using dbt, Python, and SQL, all executed directly within the Snowflake platform.

DataOps.live solutions also include Metis, an AI-powered data engineering agent that automates code generation, documentation, and data product setup to accelerate delivery and reduce risk.

The platform integrates with leading ecosystem partners such as Fivetran, Informatica, dbt Labs, Monte Carlo, Collibra, and BigID to create an end-to-end DataOps fabric spanning ingestion, transformation, and delivery.

Representative customers: AT&T, EutelSat, The Walt Disney Company, AstraZeneca, Snowflake

Google

Google was founded in 1998 and is headquartered in Mountain View, California. It offers Cloud Composer, a fully managed Apache Airflow offering, as a key orchestration tool within a broader DataOps practices.

Cloud Composer is a fully managed workflow orchestration service for creating, scheduling, and managing pipelines across clouds and on-premises. It integrates natively with BigQuery, Dataflow, Dataproc, Datastore, Cloud Storage, Pub/Sub, and AI Platform to orchestrate end-to-end Google Cloud workloads. Built on Apache Airflow and using Python, it now offers preview support for Airflow 3.0.

By using Cloud Composer instead of a local instance of Apache Airflow, you can benefit from the best of Airflow with no installation or management overhead. Cloud Composer environments are self-contained Airflow deployments based on Google Kubernetes Engine. It provides interfaces for managing environments, Airflow instances that run within environments, and individual DAGs. Developers can create and configure Cloud Composer environments in the Google Cloud console, Google Cloud CLI, Cloud Composer API, or Terraform. Developers can manage DAGs from the Google Cloud console, native Airflow UI, or by running Google Cloud and Airflow CLI commands.

Pricing for Cloud Composer is consumption-based, as measured by vCPU/hour, GB/month, and GB transferred/month in Cloud Composer 2, or DCU/hour, GB/month, and GB transferred/month in Cloud Composer 3.

Representative customers: Deutsche Bank, Vodafone

IBM

IBM was founded in 1911 and is headquartered in Armonk, New York, U.S. Its DataOps products are watsonx.data intelligence, which is an integrated product including IBM Knowledge Catalog, IBM Manta Data Lineage, and IBM Data Product Hub, and watsonx.data integration, which includes IBM Data Observability by Databand.

In 2025, IBM launched an AI assistant that uses natural language for DataOps tasks like job scheduling and transformation explanation. It integrated Databand's observability with data integration tools — DataStage, Data Replication, and StreamSets — across batch, replication, and streaming. Users can abstract pipeline logic from execution engines for transformation reuse. A new DataOps agent prioritizes issues via context-aware impact analysis and speeds resolution with AI-driven root cause insights.

IBM continues supporting the pipeline orchestration and observability built in watsonx.data integration and in external products (dbt, Apache Airflow, Informatica, Snowpipe, Databricks), including auto-healing of data pipelines and configurable task dependencies. Pipelines can be scheduled or triggered through external automation.

DataStage and StreamSets generate configurable CI/CD unit tests for pipeline testing and deployment automation for both batch and streaming workflows. On demand, it can redeploy tested project versions to downstream environments to progress or roll back. A unified SDK for watsonx.data integration will enable automation of tasks for administrators and pipelines as code.

Representative customers: Autodesk, Lockheed Martin

iceDQ

iceDQ was founded in 2004 and is headquartered in Stamford, Connecticut, U.S. Its DataOps product is called iceDQ.

The iceDQ platform integrates regression testing directly into data workflows, standardizing validation across development, testing, and production stages. Tests can be triggered from pipelines or configured to initiate pipeline executions.iceDQ provides DataOps observability by monitoring data assets and detecting quality anomalies across databases and datasets to ensure readiness for analytics. Users can generate or custom-code validation rules to build repeatable test automation suites for dry-run pipeline testing.

Recent enhancements include AI-powered observability for autonomous anomaly detection, rule generation, and proactive monitoring. With support for low-code/no-code rule creation, multisource data comparison, and prebuilt dashboards, iceDQ enables granular exception reporting. It's 20+ CI/CD integrations, including Jenkins, Azure Pipelines, Bamboo, and Git, that allow automation of data quality checks throughout the development life cycle.

The platform supports rule versioning, audit trails, and metadata management to ensure traceability and governance across data quality processes. iceDQ also enables scheduling and dependency management for test execution and supports a hybrid deployment model, including on-premises, cloud, and containerized environments.

Representative customers: Albertsons Companies, Fidelity, Morgan Stanley, Pfizer, S&P Global

lakeFS

lakeFS was founded in 2020 and is headquartered in New York, NY. Its DataOps product is called lakeFS, and its managed service version is known as lakeFS Cloud.

lakeFS is a data version control platform that enables Git-like operations with data lakes. Its core capabilities are open-source and also part of a comprehensive enterprise platform that is sold commercially.

lakeFS enables branching, committing, and reverting changes to multimodal data stored in object storage systems like Amazon S3, Azure Blob, and Google Cloud Storage, and any S3-compatible storage. While it does not orchestrate pipelines directly, it integrates with external orchestration tools like Apache Airflow, Dagster, and CI/CD platforms natively as well as via APIs and pre-commit/post-commit hooks. These integrations enforce data quality checks and automate workflows around data life cycle events.

lakeFS supports both batch and streaming workflows by enabling zero-copy branching for isolated environments. This allows safe testing without data duplication or production impact. Its observability features — commit history, rollback, and audit trails — ensure reproducibility and governance.

Infrastructure-as-code is supported via CLI and SDKs. LakeFS enables collaborative development, reproducible experiments, and data development life cycle management. Its versioning model supports scalable, governed data operations and is increasingly adopted in enterprises seeking to modernize their data lake architecture or build their AI infrastructure.

Representative customers: Lockheed Martin, U.S. Department of Energy, NASA, Volvo, Microsoft

Microsoft

Microsoft was founded in 1975 and is headquartered in Redmond, Washington. It offers Apache Airflow Job (AAJ) in Microsoft Fabric as its current DataOps product.

Apache Airflow Job (AAJ) in Microsoft Fabric is the next generation of Workflow Orchestration Manager (WOM) in Azure Data Factory.

AAJ in Fabric is a fully managed service for authoring and running Python-based DAGs. It has several workload management features, including autoscaling, suspending idle operators, pausing and resuming service, and high availability. AAJ can also run Fabric pipelines and notebooks. In addition to AAJ, Microsoft Fabric natively supports orchestration through Data Factory pipelines and built-in monitoring via Fabric's observability layer. These capabilities allow users to orchestrate data movement, transformation, and workloads across domains.

WOM in Azure Data Factory is a managed service for Apache Airflow on the Azure cloud. It offers Python-based DAGs for defining the data orchestration process. WOM provides workspace setup for Apache Airflow projects and native integration with other Azure services. Developers benefit from streamlined Airflow upgrades and patches, as well as workflow monitoring. Fabric also enhances WOM's orchestration capabilities by enabling interoperability with Power BI, Synapse Real-Time Intelligence, and Spark-based notebooks, supporting end-to-end data and analytics workflows.

Representative customers: Unavailable

Prefect

Prefect was founded in 2018 and is headquartered in Washington, D.C., U.S. Its DataOps product is Prefect Cloud, as well as an open source option.

Prefect Cloud is primarily a data pipeline orchestration tool. It orchestrates pipelines built in Databricks, Snowpipe, dbt, and Fivetran, to name a few. It treats jobs (ETL jobs, application integration jobs, robotic process automation [RPA] jobs) as workflows and executes these workflows as Python code.

Its pipeline orchestration can be parameterized to invoke the same workflow on different paths depending on runtime conditions. Integration with event-based systems like Kafka and RabbitMQ enables workflows to be triggered from external systems.

Prefect focuses on workflow execution observability, including the ability to track, visualize, and audit. For data observability, the company partners with vendors such as Monte Carlo and Bigeye. Prefect offers an open-source framework called Marvin AI that generates error summaries on metrics like lateness, success rate, and job duration and notifies Slack or Microsoft Teams.

For environment management, data infrastructure can be dynamically provisioned and configured using Python code and YAML, which support infrastructure-as-code.

In 2025, Prefect introduced ControlFlow, a new Python framework for building agentic workflows. Also, following up on the high demand for the community edition of its FastMCP project, Prefect has launched FastMCP Cloud for data and AI engineering orchestration.

Representative customers: Cash App, dbt Labs, Endpoint, Cox Automotive, Rent The Runway

Stonebranch

Stonebranch was founded in 1999 and is headquartered in Atlanta, Georgia, U.S. It offers two DataOps products: Universal Automation Center (UAC) and Universal Data Mover Gateway (UDMG).

UAC orchestrates data pipelines across hybrid environments — on-premises, cloud, and containerized systems — using a central controller, APIs, integrations, webhooks, and agents. Developers and data automation professionals build workflows using a visual designer, with support for infrastructure-as-code, environment templating, and jobs-as-code.

UAC supports time-based and event-based triggers, enabling pipelines to respond to system conditions. It monitors execution to support SLA tracking, root-cause analysis, and auditability.

As part of their data pipeline orchestration offering, customers use UAC to manage pipeline DataOps stages, transfer data from external sources, provision infrastructure, manage credentials, and automate container functions. It integrates with Kubernetes, Terraform, AWS CloudFormation, and other cloud infrastructure tools.

UAC supports pipeline observability through dashboards and OpenTelemetry integration, enabling real-time monitoring and diagnostics. Simulation capabilities allow test automation before deployment. Environment management is supported via infrastructure-as-code provisioning and self-service automation.

UAC primarily orchestrates and observes data pipelines built external to its platform. Its deployment model covers on-premises, single/multiple clouds, hybrid, SaaS, and container-based functions.

Representative customers: Not available

Unravel Data

Unravel Data was founded in 2013 and is headquartered in Mountain View, California, U.S. Its DataOps products are named Unravel for Databricks, Unravel for Snowflake, Unravel for Big Query, Unravel for Amazon EMR, and Unravel for Cloudera.

Unravel Data supports orchestration and observability capabilities on Databricks, Snowflake, Google BigQuery, Amazon EMR, and Cloudera. It interfaces with over 50 open-source, cloud, and commercial services. A core focus of Unravel is to combine data pipeline orchestration, observability, environment management, and FinOps with AI to optimize DataOps life cycle performance, platforms, and costs.

Unravel introduced agentic AI capabilities, including specialized agents for FinOps, DataOps, and data engineering. These agents automate operations cost governance, root cause analysis, performance tuning, and code reviews with supporting autonomous remediation.

Unravel Data's orchestration capabilities include dependency bottleneck identification, scheduling, monitoring pipelines/workflows, impact analysis, and alerting. Unravel interacts exclusively with data pipelines built externally outside of the tool. Additionally, CI/CD DevOps integration, "dry run" pipeline testing, and lineage tracing are incorporated.

The data pipeline observability component uses AI to correlate pipeline and environmental measurements, generate AI-assisted recommendations, and automate remediation to reduce costs, optimize instance parameters, and improve code and configuration settings across pipeline execution control planes. Unravel integrates data quality checks from external tools into its observability platform, then uses AI to enhance contextual insights and automated remediation.

Representative customers: Humana, Intel, A.P. Moller — Maersk, Novartis, Wells Fargo

Market Recommendations

D&A leaders responsible for data management operations and looking for suitable DataOps tools should:

- Inventory core data platforms, cloud service providers (CSPs), and integration patterns. Document existing workflows to identify operational gaps. Confirm compatibility with databases, APIs, and emerging protocols like Model Context Protocol (MCP) and agent-to-agent (A2A) to ensure future-ready AI capabilities and intelligent agentic coordination.

- Evaluate AI-assisted DataOps tools for their ability to accelerate workflow integration, task automation, and pipeline reliability. Prioritize solutions that support CI/CD pipelines, environment management, and end-to-end observability to reduce deployment friction. Look for embedded AI and agentic features such as autonomous monitoring, anomaly detection, and optimization recommendations, along with NLP interfaces to minimize manual effort.

- Select tools that orchestrate, observe, and automate across siloed environments, enabling consistent monitoring, implementation, and governance. Ensure unified dashboards for lineage, SLAs, and health checks. Account for overlapping capabilities from Data Management Platforms (DMPs) to avoid duplication.

- Map overlaps across DataOps, data observability, DMPs, and data integration tools to eliminate capability sprawl. Favor modular, interoperable solutions that integrate with your existing ecosystem and support open standards, reducing complexity while maintaining flexibility for future analytics and AI requirements.

- Align tool selection to organizational maturity and use cases. For early-stage teams or those focused on traditional batch ETL, prioritize tools with strong orchestration and test automation. For advanced teams pursuing real-time analytics or AI-ready data, evaluate platforms with agentic AI, autonomous orchestration, and support for streaming, semantic enrichment, and model life cycle integration.

- Match features to stakeholder priorities. For data consumers, emphasize faster cycle times, SLA adherence, and quality enforcement. For data managers, focus on failure prediction, troubleshooting efficiency, and robust schema/pipeline change management and orchestration.

Evidence

The findings and vendors included in this research draw on:

- Gartner client interactions.
- Vendor briefings in which vendors provided Gartner with updates on their strategy, market positioning, recent key developments, and product roadmaps
- Analysis of end-user interactions about DataOps from January 2024 through August 2025.

2024 Gartner Evolution of Data Management Survey. This survey was conducted to establish the characteristics of a successful data management function and understand the future operating model, architecture, and investment areas of data management teams. It also sought to identify what makes data management leaders successful in delivering data to business domains, meeting their SLAs, and being able to defend their position by showcasing value. The research was conducted online from August through September 2024 among 248 respondents from across the world. Respondents were required to have involvement in, knowledge of, and responsibility for implementing the data management side of the D&A strategy at their organizations. Disclaimer: The results of this survey do not represent global findings or the market as a whole but reflect the sentiments of the respondents and companies surveyed.

Gartner's Initial Market Coverage

This Market Guide provides Gartner's market coverage since the last publication of the market guide and focuses on the market definition, rationale for the market, and market dynamics.

Note 1: DataOps Market Size Estimation

The two main approaches to getting an approximate market size for DataOps tools are as follows:

- The application development market size was around $24 billion at the end of 2024 (see Market The data management software market size was determined to be around $13 billion at the end of 2024 (see Market Share: Enterprise Infrastructure Software, Worldwide, 2024). This includes investment in commercial vendor software and excludes custom solutions built in-house by organizations. Market penetration of DataOps as a practice was 20 to 50% in 2025 (see Hype Cycle for Data Management, 2025); let's take 50% as an optimistic estimate, as Gartner has witnessed a significant increase in client interest for DataOps tools recently. Let us assume that since all organizations that make up the $13 billion purchased vendor tools for data management, they'll do the same for DataOps as well, as the market matures. This means that the total addressable market (TAM) for DataOps tools can be estimated to be 50% of $13 billion, which is around $6.5 billion.
- Based on various non-Gartner sources, the DevOps tools market was sized at $13B in 2024. This means that DevOps is roughly 55% of the application/software development market right now. Applying the same logic for DataOps, we can use 55% to carve out the DataOps market from the data management software market of $13 billion, which gives us a rough DataOps TAM of $7 billion.

Using these two approaches, we can assume an approximate range for the DataOps tools TAM to be between $6.5 billion and $7 billion.

Document Revision History

Market Guide for DataOps Tools - 8 August 2024

Market Guide for DataOps Tools - 5 December 2022

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Streamline Data Delivery Using 5 Essential DataOps Practices

Use These 5 DataOps Practices to Drive Collaboration and Governance

Data and Analytics Essentials: What You Need to Know About DataOps

Apply 5 Top Practices to Improve Data Engineering

Develop 3 Essential Practices for Data Management Operations

Market Guide for Data Observability Tools

Table 1: Representative Vendors in the DataOps Market

Vendor	Product Names	DataOps Category	Geographic Presence	Sample of Representative Industry Coverage
Acceldata	Data Observability Cloud	Specialist (data pipeline observability)	All regions	Financial, telco, consumer packaged goods (CPG), technology
Ascend.io	Data Automation Cloud	Generalist	North America, Europe, APAC	Media, technology, healthcare, retail, financial services
Astronomer	Astro, Astro Observe	Orchestrator	All regions	Financial services, insurance, manufacturing, healthcare, retail
Amazon Web Services (AWS)	Amazon Managed Workflows for Apache Airflow (MWAA)	Orchestrator	All regions	Advertising, automotive, manufacturing, utilities, financial services
BMC Software	Control-M, BMC Helix Control-M	Orchestrator	Americas, Europe, Asia/Pacific	Financial services, telco, retail, logistics, government
Composable Analytics	Composable DataOps Platform	Generalist	North America	Financial services, pharma, telco, utilities, government
Dagster Labs	Dagster+	Orchestrator	North America, Europe, APAC	Financial services, pharma, energy, retail, utilities
Datagaps	DataOps Suite,	Specialist	North America, Europe,	Financial services, healthcare,

	ETL Validator, BI Validator, DQ Monitor, Test Data Mgr.	(data pipeline test automation)	Australia, India	education, energy, technology, manufacturing, telco
DataKitchen	DataOps Observability, DataOps TestGen, DataOps Automation	Generalist	Americas, Europe	Utilities, financial services, life sciences, manufacturing, healthcare
DataOps.live	DataOps.live	Generalist	U.S., Europe	Communications, technology, media, healthcare, life sciences
Google	Cloud Composer	Orchestrator	All regions	Healthcare, retail, financial services, media, manufacturing
IBM	watsonx.data intelligence	Generalist	All regions	Financial services, government, insurance, industrial, healthcare
iceDQ	iceDQ	Specialist (Data Pipeline test automation)	North America, Europe, APAC	Utilities, Financial Services, government, insurance, manufacturing
lakeFS	lakeFS	Specialist (Environment Management, Test automation)	North America, Europe	Financial services, healthcare, manufacturing, utilities
Microsoft	Azure Workflow Orchestration Manager	Orchestrator	All regions	Not available

Prefect	Prefect Cloud	Orchestrator	Americas, EMEA, APAC	Financial services, insurance, energy, education
Stonebranch	Universal Automation Center (UAC), Universal Data Mover Gateway (UDMG)	Orchestrator	Americas, EMEA, APAC	Financial services, insurance, Pharma, retail, manufacturing
Unravel Data	Unravel for Databricks, Unravel for Snowflake, Unravel for BigQuery, Unravel for Amazon EMR, Unravel for Coudera	Specialist (data pipeline observability)	North America, Western Europe, APAC	Financial services, healthcare, life sciences, retail, telco

Source: Gartner