

UNLOCK YOUR DATA INITIATIVES WITH DATAOPS



Across every industry, companies continue to put increased focus on gathering data and finding innovative ways to garner actionable insights. Organizations are willing to invest significant time and money to make that happen.

According to [IDC](#), the data and analytics software and cloud services market reached \$90 billion in 2021 and is expected to more than double by 2026 as companies continue to invest in artificial intelligence and machine learning (AI/ML) and modern data initiatives. It is worth noting that a significant amount of data storage, processing, and insights is happening in the cloud, given the elastic compute and storage capabilities available.

However, despite high levels of investment, data projects can often yield lackluster results. A recent survey of advanced major analytics programs by [McKinsey](#) found that companies spend 80 percent of their time doing repetitive tasks such as preparing data, where limited value-added work occurs. Additionally, they found that only 10 percent of companies feel they have this issue under control.

So why are data project failure rates so high despite increased investment and focus?

Many variables can impact project success. Often cited factors include project complexity and limited talent pools. Data scientists, cloud architects, and data engineers are in short supply globally. Companies are also recognizing that many of their data projects are failing because they struggle to operationalize the data initiatives at scale in production.

Unlocking data with DataOps

This has led to the emergence of DataOps as a new framework to overcoming common challenges. DataOps is the application of agile engineering and DevOps best practices to the field of data management to help organizations rapidly turn new insights into fully operationalized production deliverables that unlock business value from data.

The number of organizations adopting DataOps practices to help them unlock their data is increasing exponentially, so much so that analyst firms have started tracking DataOps tools as a market.

In 2022, industry analyst Gartner® published the [*Market Guide for DataOps Tools*](#), in which it provided this market definition:

“DataOps tools provide greater automation and agility over the full life cycle management of data pipelines in order to streamline data operations. The core capabilities of a DataOps tool include:

- **Orchestration:** Connectivity, workflow automation, lineage, scheduling, logging, troubleshooting, and alerting
- **Observability:** Monitoring live/historic workflows, insights into workflow performance and cost metrics, impact analysis
- **Environment Management:** Infrastructure as code, resource provisioning, environment repository templates, credentials management
- **Deployment Automation:** Version control, release pipelines, approvals, rollback, and recovery
- **Test Automation:** Business rules validation, test scripts management, test data management”

As the Gartner market definition indicates, orchestration of data pipelines is a key element of DataOps capabilities. However, data workflow orchestration comes with its own set of challenges.

Data orchestration challenges

Most data pipeline workflows are immensely complex and run across many disparate applications, data sources, and infrastructure technologies that need to work together. While the goal is to automate these processes in production, the reality is that without a powerful workflow orchestration platform, delivering these projects at enterprise scale can be expensive and often requires significant time spent doing manual work.

Data workflow orchestration projects have four key stages: ingestion, storage, processing, and delivering insights to make faster and smarter decisions.



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

Ingestion involves collecting data from traditional sources like enterprise resource planning (ERP) and customer resource management (CRM) solutions, financial systems, and many other systems of record in addition to data from modern sources like devices, Internet of Things (IoT) sensors, and social media.

Storage increases the complexity with numerous different tools and technologies that are part of the data pipeline. Where and how you store data depends a lot on persistence, the relative value of the data sets, the refresh rate of your analytics models, and the speed at which you can move the data to processing.

Processing has many of the same challenges. How much pure processing is needed? Is it constant or variable? Is it scheduled, event-driven, or ad hoc? How do you minimize costs? The list goes on and on.

Delivering insights requires moving the data output to analytics systems. This layer is also complex, with a growing number of tools representing the last mile in the data pipeline.

With new data and cloud technologies being frequently introduced, companies are constantly reevaluating their tech stacks. This evolving innovation creates pressure and churn that can be challenging because companies need to easily adopt new technologies and scale them in production. Ultimately, if a new data analytics service is not in production at scale, companies are not getting actionable insights or achieving value.

Achieving production at scale with the right platform

Successfully running business-critical workflows at scale in production doesn't happen by accident. The right workflow orchestration platform can help you streamline your data pipelines and get the actionable insights you need. That makes finding the right workflow orchestration platform vital.

With that in mind, here are eight essential capabilities to look for in your workflow orchestration platform:

1. **Support heterogeneous workflows:** Companies are rapidly moving to the cloud, and for the foreseeable future will have workflows across a highly complex mix of hybrid environments. For many, this will include supporting the mainframe and distributed systems across the data

center and multiple private and/or public clouds. If your orchestration platform cannot handle the diversity of applications and underlying infrastructure, you will have a highly fragmented automation strategy with many silos of automation that require cumbersome custom integrations to handle cross-platform workflow dependencies.

2. **Service level agreement (SLA) management:** Business workflows, ranging from ML models predicting risk to financial close and payment settlements, all have completion SLAs that are sometimes governed by guidelines set by regulatory agencies. Your orchestration platform must be able to understand and notify you of task failures and delays in complex workflows, and it needs to be able to map issues to broader business impacts.
3. **Error handling and notifications:** When running in production, even the best-designed workflows will have failures and delays. It is vital that the right teams are notified so that lengthy war room discussions just to figure out who needs to work on a problem can be avoided. Your orchestration platform must automatically send notifications to the right teams at the right time.
4. **Self-healing and remediation:** When teams respond to job failures within business workflows, they take corrective action, such as restarting a job, deleting a file, or flushing a cache or temp table. Your orchestration platform should enable automation engineers to configure such actions to happen automatically the next time the same problem occurs.
5. **End-to-end visibility:** Workflows execute interconnected business processes across hybrid tech stacks. Your orchestration platform should be able to clearly show the lineage of your workflows. This is integral to helping you understand the relationships between applications and the business processes they support. This is also important for change management. When making changes, it is vital to see what happens upstream and downstream from a process.
6. **Self-service user experience (UX) for multiple personas:** Workflow orchestration is a team sport with many stakeholders such as data teams, developers, operations, business process owners, and more. Each team has different use cases and preferences for how they want to interact with the orchestration tools. This means your orchestration platform must offer the right user interface (UI) and UX for each team so they can benefit from the technology.
7. **Production standards:** Running workflows in production requires adherence to standards, which means using correct naming conventions, error-handling patterns, etc. Your orchestration platform should have a mechanism that provides a very simple way to define such standards and guide users to the appropriate standards when they are building workflows.
8. **Support DevOps practices:** As companies adopt DevOps practices such as continuous integration and continuous deployment (CI/CD) pipelines, the workflow development, modification, and even infrastructure deployment of workflows, your orchestration platform should be able to fit into modern release practices.

Control-M and BMC Helix Control-M

DataOps tools and methodologies can help you make the best use of your data investment. But if you want to succeed in your DataOps journey, you must be able to operationalize the data. Control-M (self-hosted) and Helix Control-M (SaaS) provide a layer of abstraction to simplify the orchestration of complex data pipelines. These application and data workflow orchestration platforms enable end-to-end visibility and predictive SLAs across any data technology or infrastructure.



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

Control-M and Helix Control-M can help you orchestrate your data pipelines, put your data to effective use, and improve your data-driven business outcomes. Both platforms are used by thousands of companies globally and are proven to help companies run data pipeline workflows in production at scale.

Here are some examples of the robust capabilities Control-M and Helix Control-M have and how they can help you streamline your data pipeline workflow orchestration:

Robust integrations

The tools required to run a modern business vary widely. Often, each department utilizes its own technologies, requiring manual scripting to connect workflows across the business. Control-M and Helix Control-M feature a vast [library of out-of-the-box integrations](#) that allow businesses to orchestrate the latest technologies.

SLA management and impact analysis

With Control-M and Helix Control-M, you can track the status of business service levels along with corresponding workflows, so you know exactly how business services are performing at any given time. The two platforms can predict that a service will be late if a job is delayed or has failed upstream because they are using historical data to calculate how long a downstream job usually takes to run. Using this data, they can notify stakeholders not only that a particular job is late, but which business services are at risk of being delayed.

Python client

Many teams within an organization need to interact with your workflow orchestration platform for various reasons. Developers are a particularly important stakeholder in the orchestration process. They develop the applications that will run in production and be orchestrated by Control-M and Helix Control-M. The Python client allows developers to natively invoke their functions from their Python code.

Visibility for business users

Business users are an important stakeholder, as well. They are ultimately responsible for the timely delivery of the services they own. With the Control-M mobile app and web interface, they can track the status of their workflows anytime, from anywhere, without having to contact the application teams or operations for status updates.

The need for data is on the rise and shows no signs of abating, which means that having the ability to

store, process, and operationalize that data will remain crucial to the success of any organization. DataOps practices backed by the powerful data orchestration capabilities of Control-M and Helix Control-M can help you orchestrate data pipelines, streamline the data delivery process, and improve business outcomes.

1. **Market Guide for DataOps Tools*; December 5, 2022; Robert Thanaraj, Sharat Menon, Ankush Jain

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

To learn more about how Control-M and Helix Control-M can help you deliver data-driven outcomes faster, visit our website ([Control-M](#)/[Helix Control-M](#)).