Unlocking the Power of DataOps for the Enterprise
As data pipelines grow in size and complexity, DataOps continues to gain a bigger foothold in enterprises hungry for fast, actionable insights.

Faster delivery and greater operations productivity are the top benefits cited by users. At the same time, implementing an effective DataOps program requires significant technological, process, and cultural changes.

Moreover, new trends and best practices are continually evolving—from the emergence of the DataOps engineer to the rise of data mesh, data observability, and related disciplines like DataSecOps and DataGovOps.

WHAT IS DATAOPS?

DataOps—short for data operations—is a collection of practices, principles, technologies, and staffing positions meant to make sure data is handled efficiently.

The search for more agile data and analytical processes has led to the emergence of DataOps frameworks that can improve an organization’s responsiveness to the business users who demand more information.

Like its earlier cousin DevOps, the aim of DataOps is to avoid very large and complex data projects that take months or years to deliver value by breaking them up into smaller projects that can be delivered in a short amount of time.

WHERE DATAOPS IS HEADING

According to a recent market.biz report, the DataOps platform market is expected to grow from $1,198 million USD in 2022 to $3,856 million USD by 2032, at a compound annual growth rate (CAGR) of 25.7% during the forecast period. The major drivers of the DataOps platform market include the need for real-time data insights, increasing adoption of cloud-based solutions, and the rise in adoption of agile and DevOps practices.

DataOps platforms enable organizations to manage and monitor their data pipelines, ensuring that data is flowing smoothly and efficiently from collection to analysis. By automating many of the tasks associated with data management, DataOps platforms can help organizations reduce costs and improve operational efficiency.

DataOps brings people, processes, and technology together to orchestrate the effective, efficient, and secure flow of data within an enterprise. To do that, organizations must have in place key components in all three areas.

More specifically, experts say organizations that are seeking to leverage DataOps should understand the following:

• What data assets they have and the quality of that data;
• How that data currently flows through the enterprise;
• Whether data siloes continue to exist and how to eliminate them;
• How the business wants to use data;
• What data governance exists; and
• The technology components and talent they have to support all of those elements.

Additionally, companies should prioritize the following:

• Replace legacy data tech stacks with modern ones that provide complete visibility into the data pipeline;
• Invest in data literacy training across the enterprise; and
• Upskill their data teams, so they’re prepared to work in this new environment.
CHALLENGES TO UNDERSTANDING DATAOPS

According to a study by the IDC, “IDC’s BMC DataOps Survey,” quality, trust, and security are the top challenges to business value from DataOps investments.

DataOps can be complex and skills are often lacking, the report stated. The top reported obstacles include:
- Integration complexity
- High costs of consultants and outside experts
- Lack of staff time/skills
- Concerns over business disruption

To avoid these complications, the survey recommended creating a DataOps orchestration leader. This person is exceptionally prepared to make data-driven decisions in uncertain times. IDC reports that a company with a DataOps orchestration leader:
- Demonstrated business outcome improvements;
- Promotes data-driven, decision-making quality and resiliency;
- Incorporated integration and analytics across multiple data sources and diverse data formats;
- Committed to comprehensive, organization-wide data pipeline orchestration;
- Provided holistic, end-to-end architecture to standardize the use of distributed, diverse, and dynamic data;
- Invested in a unified, full-stack orchestration platform

DEPLOYING DATAOPS WITHIN THE ORGANIZATION

According to BARC research, in general, DataOps and MLOps as process-oriented concepts can be implemented with open source and commercial tool stacks alike. The study shows that users of commercial tools are less likely to experience overwhelming complexity as soon as ML models are deployed. The type of tool used makes little difference in the phase before the first deployment.

“This is no surprise as DataOps focuses on the realization of a manageable, maintainable, and automated flow of quality-assured data to data products whose success depends on being up to date,” said Alexander Rode, data and analytics analyst at BARC and co-author of the study. “MLOps addresses the additional special requirements of the development, deployment, and maintenance of ML models which are also data products. The goal of both concepts in this context is to achieve transparency regarding all interdependencies across involved systems along an end-to-end data pipeline and to foster collaboration between experts by making the process of developing, updating, and maintaining ML models more agile and efficient.”

According to IDC, by 2023, 60% of organizations will start implementing DataOps programs to reduce the number of data and analytics errors by 80% and to boost trust in analytics outcomes and efficiency of data-native workers. This focus on quality will improve the level of trust in data, data analytics, and data science as it helps organizations adapt to new business needs.

DataOps tools offer powerful automation and agility across the full lifecycle of data pipelines. According to Gartner, five key capabilities unite all DataOps platforms:

1. **Orchestration** capabilities include connectivity, scheduling, logging, lineage, troubleshooting, and alerting.
2. **Observability** enables monitoring live/historic workflows, insights into workflow performance and cost metrics, as well as impact analysis.
3. **Environment Management** features cover infrastructure-as-code (IaC), resource provisioning, and credential management.
4. **Deployment Automation** entails version control and approvals.
5. **Test Automation** provides validation, script management, and data management.

DATAOPS USE CASES

According to IDC’s research, effective DataOps methods in use among enterprises today include data sandboxes, version control, feedback loops, and logic and data testing.

As DataOps borrows the principles of agile and DevOps methodologies, it lends itself to data science, data visualization, and data warehousing use cases by breaking down the silos in data pipelines from ingestion to analysis and visualization. DataOps is a set of best practices enabled by platforms such as data-enabling backup platforms that help break down data silos.

Data warehousing is just one area where DataOps can make an impact. It can be applied to any business process or analytic solution that involves extracting, ingesting, cleaning, moving, storing, transforming, integrating, or aggregating data.

Other areas for DataOps include artificial intelligence (AIOps), cloud migration (CloudOps), digital transformation, and Customer 360 projects—essentially, any business activity that requires agile manipulation of complex data to support or create business applications.

Data scientists are often hamstrung by a lack of access to production data and sufficient computational processing to run their models; they are forced to work with sample data on laptop computers. Conversely, many have become dependent on data engineers to create working data sets and application engineers to deploy their models in operational environments. DataOps enables data scientists to provision temporary data sandboxes and create simple data pipelines as well as deploy models with minimal IT or engineering assistance.

Data analysts armed with self-service visualization tools have an endless appetite for data sets to feed their analytical inquiries. DataOps enables data analysts to service their own data needs within a curated data environment facilitated by a data catalog and data preparation tools. The data department creates a self-service data infrastructure that balances speed and standards and fosters a culture of governance that accelerates the delivery of data without spawning data silos.

According to Gartner, leaders should keep the following in mind when looking to improve data team performance and remediate operational concerns:

- Evaluate DataOps tools based on key capabilities such as connectivity to your data stack (existing and expected), workflow automation, and release automation.
- Prioritize tools that provide a unified view of workloads across your hybrid IT data stack, while also offering features such as lineage, cataloging, resource provisioning, and environment management.
- Focus on benefits when introducing your selected tool to data managers and consumers. Data managers will be most interested in the ability to monitor workflows and receive proactive alerts; data consumers will appreciate reduced cycle times to access data and improved data integrity.

DataOps tools are part of an emerging technology category that helps organizations streamline data delivery and improve productivity with process integrations and automations.

—Stephanie Simone
As organizations continue to adopt multiple databases to serve business application demands, the infrastructure data storage components and architecture should be carefully designed to meet performance, scalability, capacity, and availability needs. With Pliops XDP, the complexity that comes with changing infrastructure data storage components for adoption of various database technologies is eliminated.”—Tony Afshary, Global VP of Products and Marketing, Pliops