# Introducing DataOps Into Your Data Management Discipline

FOUNDATIONAL Refreshed 16 February 2021, Published 31 October 2019 - ID G00376495 - 10 min read

By Ted Friedman, Robert Thanaraj

To relieve bottlenecks and barriers in delivery of data and analytics solutions, organizations need to change the way they work. By introducing DataOps techniques in a focused manner, data and analytics leaders can affect a shift toward more rapid, flexible and reliable delivery of data pipelines.

## Overview

### Key Challenges

Data and analytics teams cannot achieve the speed and reliability of project delivery they desire because too many roles, too much complexity and constantly shifting requirements make it difficult.

In most organizations, this complexity is exacerbated by limited or inconsistent coordination across the roles involved in building, deploying and maintaining data pipelines.

Data and analytics leaders often have difficulty determining the optimal pace of change when introducing new techniques — wholesale reorienting of work processes, particularly in support of complex project delivery, can be overly disruptive.

### Recommendations

Data and analytics leaders focused on modernizing their data management strategies and solutions should:

Enable greater reliability, adaptability and speed by leveraging techniques from agile application development and deployment (DevOps) in your data and analytics work.

Enable collaboration across key roles (DBAs, data engineers, integration architects, data stewards, etc.) by including them in a common process, providing an infrastructure for shared metadata, establishing/formalizing an “operations focused” role, and

providing channels for regular communication and feedback.

Begin by introducing these capabilities with a focus on requirements definition, development and monitoring — these are the activities where collaboration, communication and consistency are most relevant and important.

Maximize the chances of successful introduction of these approaches by choosing data and analytics projects that are struggling due to lack of collaboration or are overburdened by the pace of change — these create the best opportunity to show value.

## Strategic Planning Assumption

Within the next year, the number of data and analytics experts in business units will grow at three times the rate of experts in IT departments, which will force companies to rethink their organizational models and skill sets.

## Introduction

As data and analytics teams become critical to supporting more diverse, complex and mission-critical business processes, many are challenged with scaling the work they do in delivering data to support a range of consumers and use cases. Pressure to deliver faster, with higher quality and with resiliency in the face of constant change, is causing data and analytics leaders to rethink how their teams are organized and how they work. Traditional waterfall-oriented methodologies aren’t meeting the need — the distance between requirements definition and delivery of value is too great, the time required too long, and too many critical tasks get lost or degraded across role and team silos. DataOps techniques can address these challenges through a more agile, collaborative and change-friendly approach to building and managing data pipelines.

---

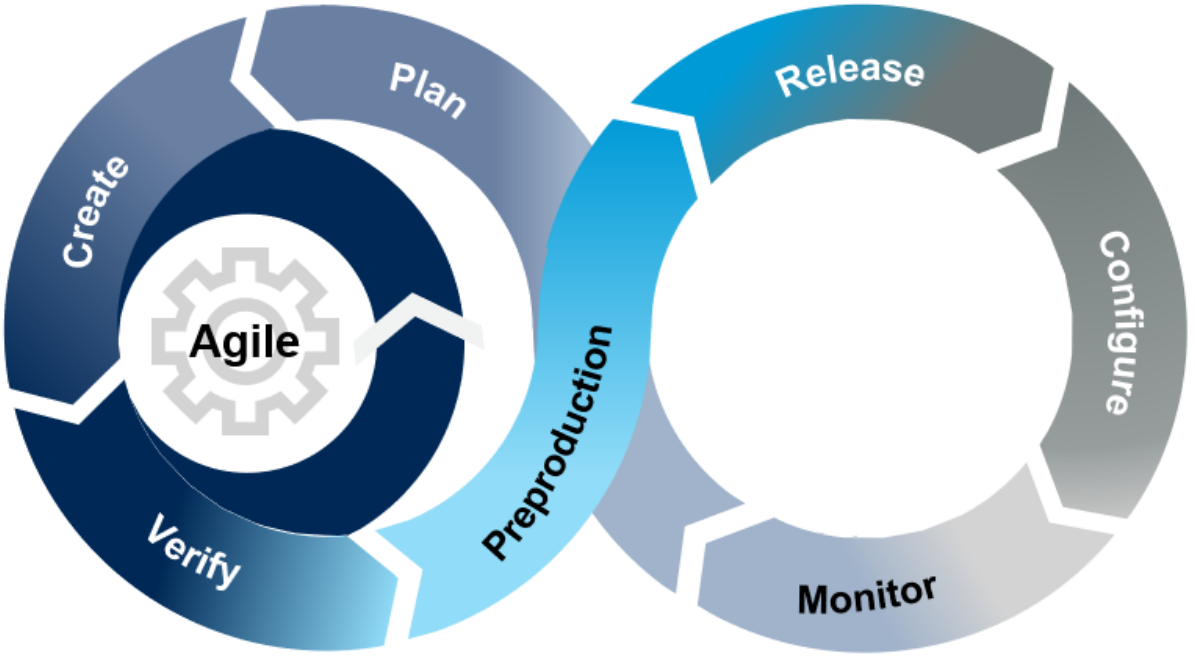
*DataOps is a collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization.*

DataOps is about a different way of working when delivering data and analytics solutions. Applying techniques adapted from the DevOps concepts (see Figure 1), which many organizations have leveraged in implementing applications, better communication and tighter collaboration results in faster deployments and greater effectiveness in reacting to change postdeployment. (See “Innovation Insight for DataOps.”)

With the increasing awareness of DataOps concepts and terminology in the market (see Evidence section), data and analytics leaders are looking for the best ways to introduce these ideas into how their teams operate. For many, DataOps represents a massive shift in approach — raising substantial change management concerns. Identifying effective and focused ways to gradually introduce DataOps techniques into data and analytics solution delivery methodologies must be a priority for leaders seeking to increase effectiveness of their distributed teams.

Figure 1. DataOps: Leveraging Learnings From the DevOps Value Chain

**DataOps: Leveraging Learnings From the DevOps Value Chain**




ID: 376495



## Analysis

### Apply Agile (DevOps) Techniques to Your Data and Analytics Work

DevOps is a set of practices intended to enable shorter system development and delivery cycles — ideally providing continuous delivery of new capabilities — with higher quality of deliverables. (See "DevOps — Eight Simple Steps to Get It Right.") These same goals resonate more and more with data and analytics leaders, their teams and the business leaders they serve. Gartner primary research data shows that speed and reliability of getting data-focused capabilities into production remain a challenge for many

organizations. Thus, it makes sense for these leaders to learn and apply DevOps concepts to their work of building and delivering data pipelines.

Specifically, data and analytics teams need to begin implementing key techniques that have delivered significant value for organizations implementing DevOps in support of application delivery efforts:

- Increased deployment frequency — shifting toward a methodology where rapid and continuous delivery of new functionality enables the organization to keep pace with change
- Automated testing — removing bottlenecks arising from traditional, human-intensive testing approaches enables higher quality amid more frequent delivery of new capabilities
- Consistent metadata and version control — closely tracking changes and communicating them rapidly across all participants in data pipeline delivery ensures more complete change management and avoids disruptions
- Monitoring — constantly tracking behavior and usage of pipeline capabilities enables more rapid identification of both flaws to be corrected and opportunities for new capabilities.
- Collaboration across all stakeholders — clear, constant communication and common metadata, particularly when changes are being introduced into a pipeline, is essential to speed of delivery.

*Recommendations:*

- Seek out teams applying DevOps practices in application development and deployment — observe how they work and talk to them about how their approaches can apply to data and analytics projects.
- Pilot DataOps principles by applying the core DevOps approaches — specifically, increasing deployment frequency, collaboration of key roles and constant monitoring — to get familiar with the approach and demonstrate the value.
- Modernize the supporting tools, especially in the areas of metadata management and testing automation, to maximize benefits.

## Enable Collaboration Across Key Roles Involved in Delivery of Data Pipelines

The modern world of data management and analytics is much more complex and diverse. Given the nature of many contemporary data and analytics use cases, the roles involved in successful project delivery are more numerous, more diverse and more distributed. Requirements ranging from fast-moving data science teams working on emerging hypotheses and requirements, to AI-driven autonomous processes, to more structured and well-defined analytics, needs to involve people inside and outside IT, across multiple business functions, and with various skills. A large number of roles are often involved:

- Business analysts
- Data architects
- Data modelers
- Data engineers
- Data stewards
- Data scientists
- Data product managers
- Machine learning (ML) model developers
- Integration architects and developers
- Application developers and administrators
- Database administrators
- Infrastructure and operations personnel
- Data security teams
- Data privacy officers

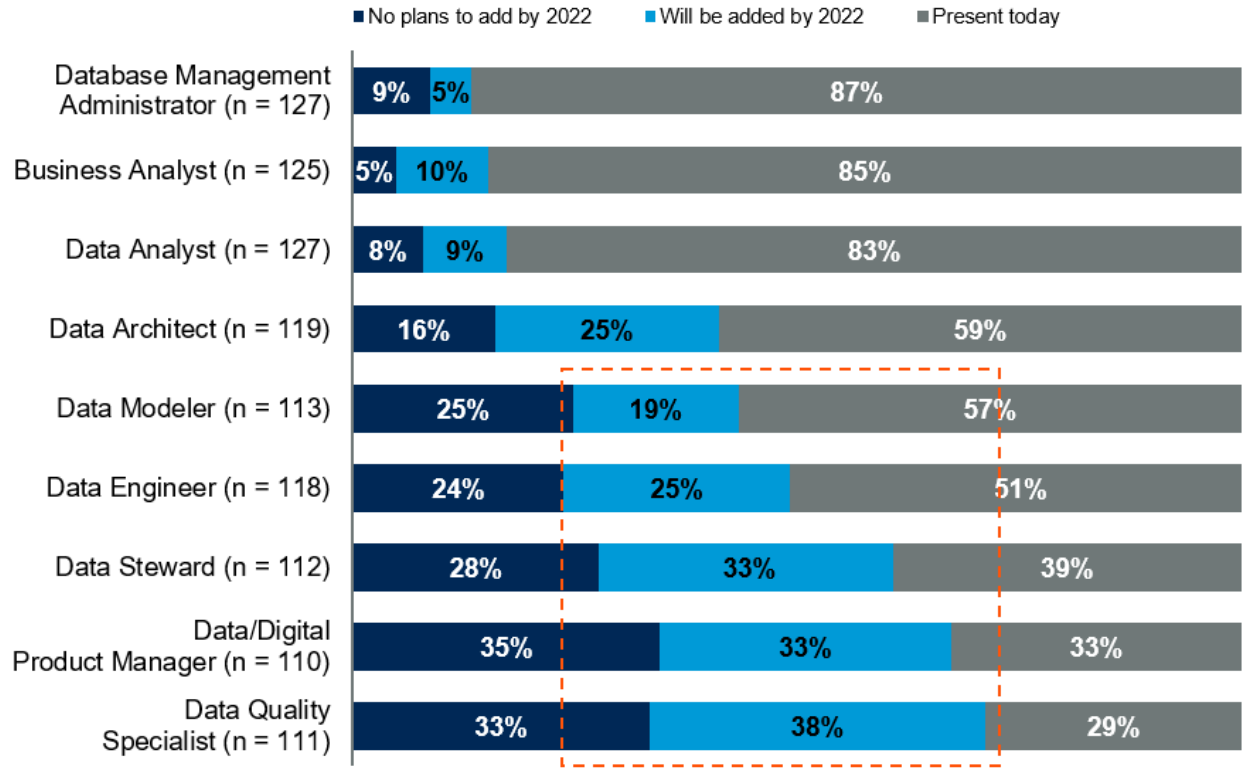
Complex data pipelines can involve all of these roles (and more) — and lack of consistent communication and coordination across them adds time and causes disconnects, which introduce quality issues. Changing business requirements often require a complete rework of all pipeline components since there is no clear visibility to dependencies and impacts across the artifacts produced by these various roles.

Many organizations are quickly adding more strategic roles that must be engaged in this work — data engineers, product managers, data stewards, etc. (See Figure 2.) A DataOps approach demands clear communication and collaboration across these roles. Each of them needs to understand what others expect and produce, and they must collectively have a shared understanding of the requirements and goals of the pipelines they are creating and evolving. Creating channels through which these roles can communicate and work together — collaboration tools, a common metadata management solution and shared metrics — is a starting point.

Figure 2. Organizations Are Rapidly Adding Key Roles Involved in Modern Data Pipeline Delivery

## Organizations Are Rapidly Adding Key Roles Involved in Modern Data Pipeline Delivery

Percentage of Respondents




Base: Gartner Research Circle Members; Excludes 'Not Sure'
Q: Which roles are currently present in your organization's data management team and which new roles do you expect to add by 2022?
Source: 2019 Gartner Research Circle Data Management Drivers
ID: 376495

Although the agile techniques expressed in DevOps rely on collective direction-setting, introduction of DataOps approaches may initially benefit from someone actively pushing the various roles to collaborate, ensuring communication is happening, and mentoring people in accepting this change. Identifying someone to play the role of “operations lead” can help to ensure early DataOps efforts show success and the concepts “stick.”

*Recommendations:*

Identify an “operations leader” — in effect, a COO or director of supply chain — for key data delivery projects. This may be a new dedicated role in the organization, or it may be fulfilled by an existing role in the ecosystem such as a senior data engineer.

Develop venues in which the key roles involved in data pipeline delivery can communicate and collaborate — both in real world (live meetings) and digitally (via collaboration tools).

Place substantial emphasis on making metadata about artifacts contributing to data pipeline delivery accessible to all relevant roles — leverage existing metadata management tooling with strength in cataloging, lineage and impact analysis, or augment as needed.

## Start With a Focus on Requirements Definition, Development and Tracking Value

To avoid introducing too much change too quickly, data and analytics leaders can focus on a subset of the steps in the value chain, rather than immediately introducing new approaches in every step executed by every role. In this way, a subset of roles and artifacts involved in data pipeline delivery can be used to pilot the relevant aspects of DataOps, rather than expecting a large and complex set of roles to change behavior all at once.

By focusing on steps that have a direct connection to business stakeholders and can also be measured, the advantages of DataOps techniques can be more clearly seen and proven:

Plan — can involve business analysts, data scientists, data stewards and data engineers, focusing on the desired business outcome.

Create — involves data engineers, integration architects and developers, data scientists and application and database administrators, producing the core deliverables that make the data pipeline work.

Monitor — potentially involves all roles, since all components of the pipeline require tracking of behavior, as well as measurement of the business value/outcome driving the need for the project from the beginning.

*Recommendations:*

Pick a subset of the steps in the DataOps value chain to engage key roles in adopting the new approach. Plan (requirements definition), create (development) and monitor (tracking usage and value) represent ideal starting points as they are firmly tied to project goals and involve multiple roles.

Introduce agile techniques (rapid, focused and iterative) to produce the output from each of these steps to show the value of the new ways of working.

### Apply DataOps Techniques to Struggling or Failing Projects and Pipelines

For many organizations, culture and inertia present barriers to introducing change. Traditional approaches to defining and executing data-oriented projects can be so ingrained it is difficult to find an entry point for new ways of working. Rather than trying to organize an entirely new effort around DataOps techniques, advocating for a new approach on struggling efforts can be more effective.

Data and analytics leaders can maximize chances of successful introduction of these approaches by choosing data and analytics projects that are high priority but are struggling due to lack of collaboration or are overburdened by pace of change. These scenarios create the best opportunity to introduce change and show rapid value. In particular, they should seek efforts where the number of roles and teams involved is causing quality flaws and missed delivery targets which are degrading overall outcome value.

*Recommendations:*

Tap into key business stakeholders and the various teams and roles working on data-related projects — gain a sense of the progress, degree of success to date and pain points of each.

Identify well-defined and high-priority use cases for which the organization is struggling to deliver data effectively — these cases represent opportunities for breakthrough value by introducing new DataOps approaches.

Engage stakeholders and key roles to shift to a more agile and collaborative approach in key components of the project, positioning DataOps techniques as a way to break through bottlenecks the effort is experiencing.

## Evidence

Gartner end-user client inquiry data shows a 200% YoY increase in 2019 YTD on DataOps-related questions — while starting from a relatively small base, the number of organizations considering and acting on DataOps principles is rapidly growing.

Gartner's Data Management Drivers survey was conducted via an online survey from 19 August through 4 September 2019 with Gartner Research Circle Members — a Gartner-managed panel. There were 129 respondents for this survey. The survey was developed collaboratively by a team of Gartner analysts and was reviewed, tested, and administered by Gartner's Research Data and Analytics team.