

Enabling DataOps for Analytics



- Issue 1
- 2 Enabling DataOps for Analytics
- 3 Research from Gartner
Innovation Insight for DataOps

Enabling DataOps for Analytics

Modern enterprises need to quickly deliver the right data to a growing data consumer audience to drive strategic initiatives, often encompassing data science and machine learning, and thereby create competitive advantage. But many of these projects are failing because yesterday's processes and systems can no longer meet today's analytics requirements. Traditional data pipelines are breaking, and data quality is suffering.

We know well that data consumers' expectations are rising. Analytics within the lines of business are demanding ever-higher volumes, variety and velocity of data, as well as rapid data transformation for analytics. Their SLAs are increasingly difficult to meet. Data managers within IT, meanwhile, are struggling with legacy systems and processes that were built for longer, batch-oriented cycle times. These two groups tend to speak different languages, further complicating efforts to collaborate.

DataOps seeks to fix these imbalances and put data-driven initiatives back on a sustainable footing. This emerging discipline encompasses processes and technologies that improve the speed, efficiency and flexibility of data pipelines. It incorporates agile development methodology, rapid response to user feedback and continuous data integration. Picture the lean manufacturing process, with data as the product.

The following Gartner report will inform your understanding of DataOps and where to apply it. Authors Nick Heudecker, Ted Friedman and Alan Dayley correctly note that, "Most DataOps use cases will likely target traditional extraction, transformation and loading (ETL) and data integration tasks in an attempt to make these processes more reliable." We see in this report that Gartner is on the same page as us – the primary initial use cases center on data extraction, transformation, and loading (ETL) and data integration. Enter the Qlik Data Integration Platform. We partner closely with the world's largest organizations, including half the Fortune 100, to fully harness the rich potential of large, highly-varied and high-velocity datasets.

With the Qlik Data Integration Platform, enterprises can realize the promise of DataOps for analytics.

- We provide agile cloud migration by fully automating the replication and synchronization of structured data and metadata across hybrid, cloud and multi-cloud environments. We optimize data transfer over the wide area network with encrypted multipathing, enabling zero-downtime migrations and ongoing real-time updates. Our real-time change data capture software minimizes source production impact and integrates with the industry's broadest set of structured data sources and targets.
- We automate and accelerate the transformation of data for analytics. We auto-merge CDC change streams into data lake stores, on premises or in the cloud, then create, format and continuously update analytics-ready datasets – no hand scripting required. We also automate the data warehouse lifecycle, including modeling, DW creation and updates, greatly reducing your dependence on skilled developers.
- Finally, the Qlik Data Integration Platform speeds up how you catalog, manage, prepare, and deliver your trustworthy, actionable data to business users across your enterprise. You gain a secure, enterprise-scale repository of all the data your business has available for analytics, giving your data consumers a single, go-to catalog to find, understand, and gain insights from any underlying enterprise data source.

Together these solutions provide a comprehensive data integration platform to enable modern data architectures and deliver on the promise of DataOps for Analytics.

While enterprises are in the early stages of adoption, there is broad recognition of the problems driving the need for DataOps. Qlik is optimistic about the near-term upside for enterprises, based on initial successes among many of our customers. We see IT teams educating themselves on the right collaborative data management approach, making data seamlessly and continuously available, with faster initial delivery and rapid improvement cycles. DataOps will continue to gain importance as part of digital transformation initiatives, empowering IT leaders to focus more on improving communication, integration and automation of data flows enterprise-wide.

Research from Gartner

Innovation Insight for DataOps

The DataOps concept has emerged as a reaction to challenges in deployment and enhancement of data and analytics projects across the enterprise. Data and analytics leaders can use DataOps to drive organizational change and predictability for using data without massive investment.

Key Findings

- DataOps applies the traditional DevOps concepts of agility, continuous integration and deployment, and end-user feedback to data and analytics efforts. However, DataOps is not strictly a technical competency. The real focus and benefit of DataOps is as a lever for organizational change, to steer behavior and enable agility.
- DataOps is an early concept in data management. While an increasing number of vendors are positioning themselves around this concept, end-user organizations are in early pilot stages, with only a few having implemented DataOps approaches.

Recommendations

Data and analytics leaders focused on enabling rapid and agile data management solutions with DataOps must:

- Spotlight data and analytics projects that either didn't make it into production, or that failed to deliver on expected business benefits.
- Identify the business value associated with introducing DataOps by identifying the data-centric influences that led to failure.
- Identify bottlenecks, delays and breakage points where components of the environment vary, and illustrate how these are mitigated through the kind of change management that is part of the DataOps process.
- Enlist personnel to become DataOps champions by looking for staff that are already involved with diverse groups within the organization and are effective in multiple roles. These are typically the influencers that act as informal and transient hubs of information and communication.

Analysis

Data and analytics solutions are struggling to keep pace with the increasing complexity, speed of change and general growth in demand for access and delivery of diverse datasets. Teams charged with building data pipelines to ingest, combine, prepare and deliver data for various use cases (both analytical and operational) find it difficult to do so quickly and reliably.

DataOps can be understood better when doing a loose comparison to practices implemented through DevOps efforts as undertaken between application developers and IT management staff. Like DevOps, DataOps represents a change in culture that focuses on improving collaboration and accelerating service delivery by adopting lean or iterative practices where appropriate. Unlike DevOps, which focuses on operations and development teams, DataOps focuses on data managers and data consumers. These data consumers can be anyone — or anything — including IoT devices, analytical models or even people.

DevOps came about as a response to organizational friction in developing and deploying applications. DataOps is emerging, albeit early, as a response to similar frictions around the consumption and use of data across the organization.

Inevitably organizations that go down the DataOps path will look for best practices and implementation guidelines. These best practices will be detailed in future Gartner research as more successes appear within organizations. For the time being, this research is focused on a high-level description of DataOps, and is meant to be used as a tool to decide whether or not DataOps should be further explored to reduce data and analytics friction.

These similarities allow us to use much of the work developed as part of Gartner's DevOps research. Notably, the DevOps toolchain (see Figure 1) will need to be repurposed for DataOps. Today's data professionals tend to focus on the "plan" and "create" links in the toolchain, with other links either ignored or passed off to other, disconnected, groups within the organization. For DataOps to

become a reality, implementing this toolchain will be essential. But as we explain later, it won't be the only element necessary to guarantee success.

Definition

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. The goal of DataOps is to deliver value faster by creating predictable delivery and change management of data, data models and related artifacts. DataOps uses technology to automate the design, deployment and management of data delivery with the appropriate levels of governance and metadata to improve the use and value of data in a dynamic environment.

Description

The central challenge of data and analytics is getting new insights — and the data those insights require — into production. New analytics efforts face uncertainty at every step, from ideation to delivery, and that uncertainty impacts how value is realized.

Based on Gartner's 2017 Data and Analytics Trends Survey there are multiple reasons why deployment and support are challenges for end-user organizations (see Figure 2).

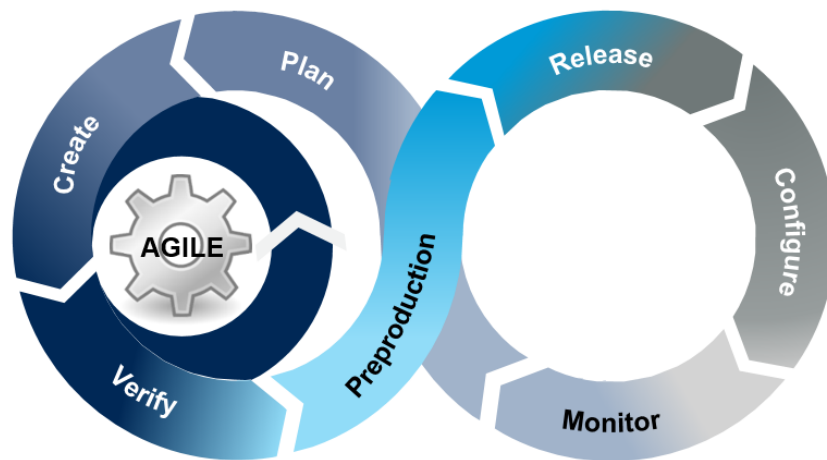
In addition to initially deploying projects, the organizational complexity (as well as politics and the evolving data infrastructure environment) creates challenges in keeping a deployed project running reliably in the face of constant change. Modifications to data structures by one system are often not communicated to downstream consumers of that data, leading to other parts of the data workflow breaking. And when changes are successfully implemented, the reasons, specifics and results are often not reflected in the associated metadata and documentation. As a result, confidence and value for data consumers are degraded.

Benefits and Uses

The nascent discipline of DataOps is gaining traction due to the desire to reduce the friction between data and analytics teams and the business units relying on their efforts. As

FIGURE 1 The DevOps — and Now DataOps — Toolchain

The DevOps Toolchain



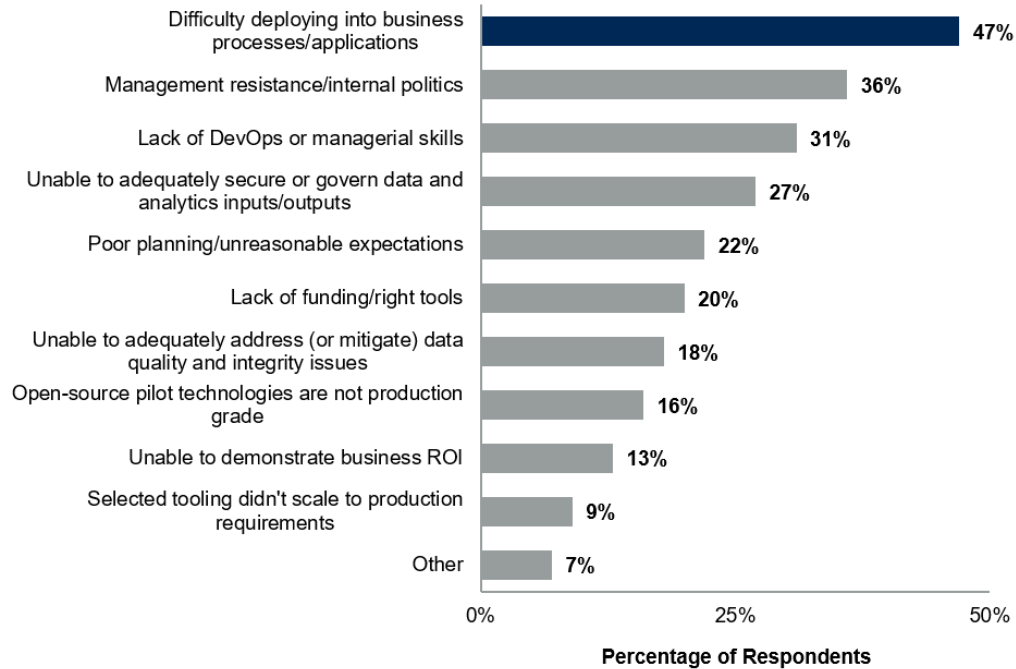
ID: 377605

© 2018 Gartner, Inc.

Source: Gartner (December 2018)

FIGURE 2 Challenges Getting Data and Analytics Projects Into Production

Barriers to Moving Projects Into Production



Base: n = 45 Gartner Research Circle members/external sample. Excludes "not sure" responses.

Question asked: Please identify your organization's specific barriers to moving projects into production. Multiple responses were allowed.

ID: 377605

© 2018 Gartner, Inc.

Source: Gartner (December 2018)

represented in Figure 2, friction may occur at multiple steps in the deployment journey, from demonstrating ROI, to integrating with existing processes, or data quality challenges.

Some examples of DataOps in practice might include formalizing the detection of changing data streams entering analytics workflows before they trigger downstream errors, or conducting A/B testing of different data sources on analytical models. However, most DataOps use cases will likely target traditional extraction, transformation and loading (ETL) and data integration tasks in an attempt to make these processes more reliable.

Beyond the data movement, integration and transformation use cases, DataOps may give enterprises an opportunity to expand how they think about data and analytics by opening the

discussion to a broader audience (such as master data management). Ultimately, practices like DevOps and DataOps are about altering how different parts of an organization collaborate. Realizing the benefits of DataOps requires a different type of collaboration than DevOps because of the different audiences involved. With DevOps, both audiences — infrastructure and operation, and application development — are inherently technical. Even though they have different interests, both groups use a common language. When DataOps is expanded outside of data engineers and data scientists to the broader business, no such common language exists. Because of this mismatch between technical and nontechnical participants, things like data literacy (see Note 1) and sharing a common language become essential.

DataOps will accrue benefits for technical audiences first, because that's the most obvious place to apply these nascent practices. Driving real organizational and cultural change around data and analytics will require a new way for disparate parties to apply context to data. Organizations should not expect to pilot DataOps and then roll it out as a full discipline with all the answers in place. Instead it will need to be an iterative approach that continually measures willful adoption, acceptable value and risks, and (most importantly) if corporate goals are being met. During the first iteration, and as political friendliness is being ascertained, repeatable and measurable metrics need to be identified. In addition, marketing the benefits of DataOps throughout the affected organization is critical for success (see Figure 3).

Adoption Rate

The current adoption rate of DataOps is estimated at less than 1% of the addressable market. There are no standards or best practices for DataOps at the present time. Some early organizations are applying DevOps principles — like continuous integration and testing — to their data pipelines, but we have yet to see anything that goes beyond this initial phase. Other organizations are simply calling what they're already doing "DataOps," hoping to draft off of the hype.

Risks

Like any new concept, DataOps comes with risks for end-user organizations. These risks are both technical and organizational in nature:

Internal DataOps washing: It is common for technologists to hear about a practice or technology implemented by another company and claim that your company can accrue the same benefits simply by purchasing the same technologies (often described as a "cargo cult").¹ However, without linking the DataOps concept to business users and outcomes, it's unlikely you will benefit. DataOps requires a change in organizational culture. You can't throw technology at the perceived problem.

Lack of defined standards and frameworks: DataOps is a new concept. There are no defined frameworks or standards, making it difficult to know if you're "doing DataOps" at all. This will

likely lead to multiple false starts as organizations figure out what DataOps means for them. A lack of early success may dissuade organizations from pursuing the beneficial cultural and collaborative changes that DataOps may bring.

DataOps washing: While multiple enterprise software vendors are talking about DataOps, their scope is largely limited to the products they sell. Technology is a core part of DataOps, but it is not the only part. DataOps is a practice, not a technology or tool. You cannot buy DataOps.

Pace of adoption and outlook: It isn't yet clear if DataOps will become a lever for broad organizational change by introducing new collaboration and communication models, or if it will remain technically focused around continuous integration and testing.

Recommendations

Data and analytics leaders focused on enabling rapid and agile data management solutions with DataOps must:

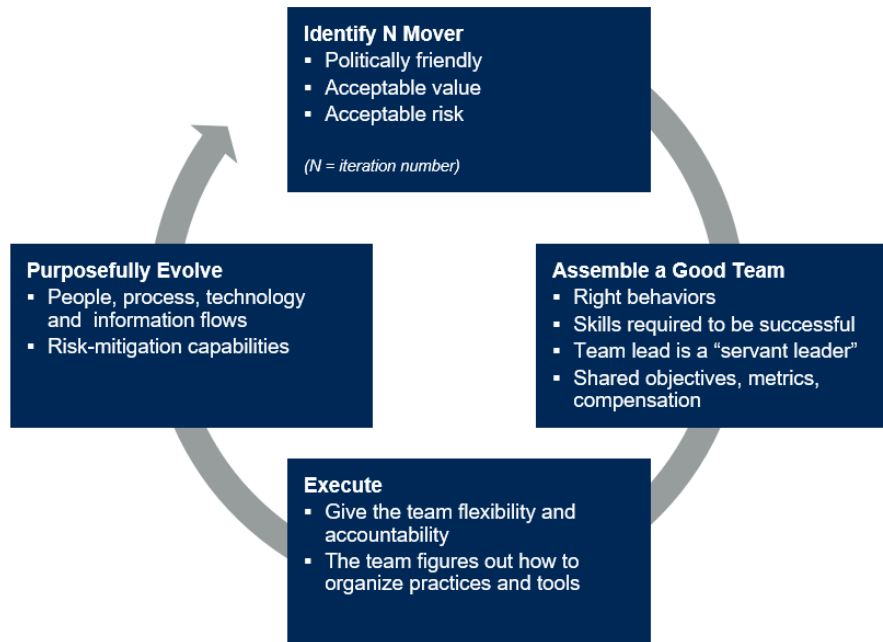
- Spotlight data and analytics projects that either didn't make it into production, or failed to deliver on expected business benefits, and then identify business value associated with introducing DataOps by identifying the data-centric influences that led to the failure.
- Identify bottlenecks, delays and breakage points where components of the environment vary, and illustrate how these are mitigated through change management that is part of the DataOps process.
- Identify and enlist personnel to become DataOps champions by looking for staff that are already involved with diverse groups within the organization, and are effective in multiple roles. These are typically the influencers that act as informal and transient hubs of information and communication.

Evidence

Gartner's 2017 Data and Analytics Trends Survey was conducted online from 23 May to 26 June 2017 among Gartner Research Circle Members — a Gartner-managed panel composed of IT and business leaders — as well as an external sample source.

FIGURE 3 Continual Learning and Improvement

Continual Learning and Improvement



ID: 377605

© 2018 Gartner, Inc.

Source: Gartner (December 2018)

In total, 196 members completed the survey (92 Gartner Research Circle members/104 external sample). Gartner Research Circle and external sample IT and IT-business members were invited to participate.

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.

Note 1. Data Literacy

Data literacy is the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, the application and resulting value.

Source: Gartner Research Note G00377605, Nick Heudecker, Ted Friedman, Alan Dayley, 27 December 2018

¹ “Cargo Cult” Wikipedia entry