451 Research | BLACK & WHITE PAPER

DataOps Unlocks the Value of Data

COMMISSIONED BY

HITACHI
Inspire the Next

About this paper

A Black & White paper is a study based on primary research survey data that assesses the market dynamics of a key enterprise technology segment through the lens of the "on the ground" experience and opinions of real practitioners — what they are doing, and why they are doing it.

ABOUT THE AUTHOR

MATT ASLETT

RESEARCH VICE PRESIDENT

Matt Aslett is a Research Vice President with responsibility for 451 Research's Data, AI and Analytics Channel – including operational and analytic databases, Hadoop, grid/cache, stream processing, data integration, data governance, and data management, as well as data science and analytics, machine learning and AI. Matt's own primary area of focus currently includes distributed data management, data catalogs, business intelligence and analytics, data science management, and enterprise knowledge graphs.

What Is DataOps?

Enterprises have always taken decisions based on data. In recent years, however, the extent to which data and analytics is embedded in the decision-making process has accelerated as enterprises have recognized that successfully converting data into business insight can be a differentiator that delivers competitive advantages resulting in enhanced agility, operational efficiencies and new revenue streams.

Converting data into insight is easier said than done, especially when using tools and approaches that were designed to serve monolithic analytics projects based on traditional development approaches. These traditional tools and approaches are unsuitable for delivering the agility and responsiveness required by all enterprises if they are to become digital businesses and respond to – if not anticipate and preempt – changing business requirements.

The term DataOps was _coined in 2015_ to describe the need for more agile approaches to data management required for enterprises to realize the potential benefits of becoming more data-driven. The use of both the term and the concept has grown since then, as has the understanding of its ability to deliver benefits related to operational agility, analytics and governance.

Definition

DataOps was initially defined fairly loosely as representing a more collaborative and cooperative approach to data management. As a result, the term can mean different things to different people in different roles in different organizations, depending on how they choose to implement collaborative and cooperative data management.

For the purposes of this report, and the State of DataOps survey on which it is based, 451 Research defines DataOps as:

> "The alignment of people, processes and technology to enable more agile and automated approaches to data management."

While the term DataOps is still in relative infancy, 451 Research has noted that many of the key concepts behind DataOps are already widely known and adopted. In fact, your organization may already be practicing and benefiting from many of these concepts without calling them DataOps.

The State of DataOps survey results clearly illustrate that there is growing interest in the concepts related to DataOps: an almost unprecedented 100% of respondents said they are currently planning (28%) or actively pursuing (72%) initiatives to deliver more agile and automated data management. This result is especially significant since we did not deliberately seek or filter for respondents that were aware of or interested in the term DataOps.

Source: 451 Research survey, commissioned by Hitachi Vantara. n=300
Q: Which of the following statements best characterizes your initiatives to deliver more agile and automated data
*   management?*

100%

of enterprises are **currently planning or pursuing initiatives to
deliver more agile and automated data management**

72%	28%
Actively pursuing	Planning
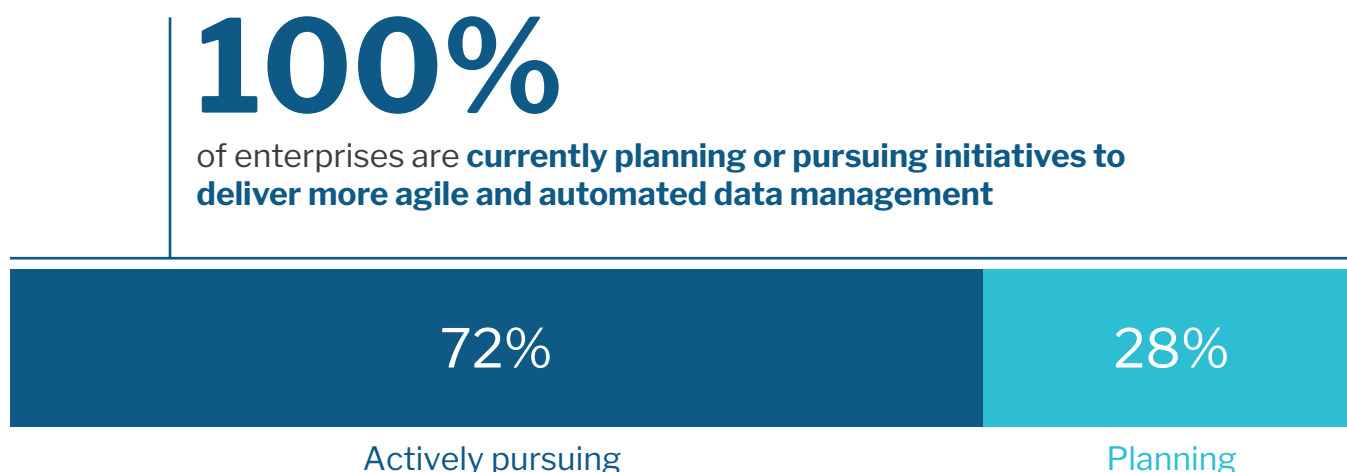
It is also worth noting that this interest in more agile and automated approaches to data management is not necessarily driven by the functional limitations of current technologies. In fact, 83% of respondents agreed with the statement: "My business or organization has the right tools to be able to manage all its data." As such, it is clear that having the right tools is not enough to deliver the agility required to support changing business requirements.

Improving the alignment of people, processes and technology is a potential benefit of DataOps, specifically in terms of reducing data friction, which arises when the demands of data consumers (such as data analysts, developers and senior decision-makers) are not met by data operators (e.g., data management and IT professionals). Data friction is a perennial problem that has been exacerbated in recent years by the growing volume of data and increased demand for the rapid development of data and analytics projects. As the survey results illustrate, DataOps has the potential to alleviate this challenge.
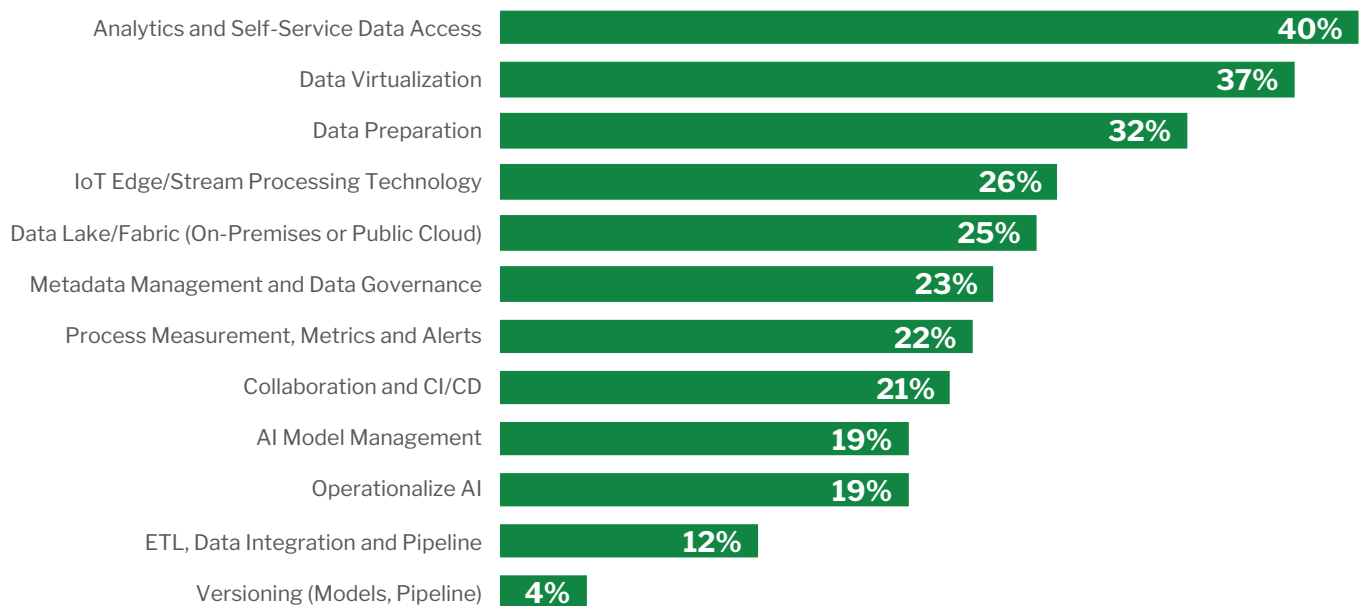
DataOps Momentum

451 Research has seen a significant increase in familiarity with the term DataOps in the past year. For example, 97% of respondents said they were familiar with the term, as defined above, and 77% of respondents indicated that they actually use the term. Additionally, it is clear that enterprises are not just talking the talk when it comes to DataOps, but taking steps to invest in its adoption and implementation. In fact, 91% of respondents indicated that they already have, or are in the process of defining, a formal DataOps strategy, while 86% of respondents reported that they plan to increase spending, investment or development related to DataOps in the next 12 months.

That investment will take the form of increased spending in a variety of technology areas, including analytics and self-service data access, data virtualization, data preparation, edge processing, data lake, and metadata management. However, it is also important to note that adoption of these technologies alone will not yield the desired results without a similar focus on people and process change.

Figure 2: Focus areas for increased DataOps-related spending
Source: 451 Research survey, commissioned by Hitachi Vantara. n=300
Q: Which of the following areas are you considering for increased spending?

Area	Percentage
Analytics and Self-Service Data Access	40%
Data Virtualization	37%
Data Preparation	32%
IoT Edge/Stream Processing Technology	26%
Data Lake/Fabric (On-Premises or Public Cloud)	25%
Metadata Management and Data Governance	23%
Process Measurement, Metrics and Alerts	22%
Collaboration and CI/CD	21%
AI Model Management	19%
Operationalize AI	19%
ETL, Data Integration and Pipeline	12%
Versioning (Models, Pipeline)	4%

Self-service data access and analytics is the most popular area for investment, and it is clearly a potential route for reducing the friction between data consumers and data operators. While IT has traditionally been the gatekeeper to analytics insight – creating reports, dashboards and data models and provisioning data warehouses and associated infrastructure in response to demand from business users – this approach is entirely unsuitable for driving real-time decision-making on large volumes of live data. In contrast, we have witnessed a shift toward self-service enablement via the democratization of data and associated infrastructure that bypasses a dependency on IT that has the risk of becoming a bottleneck to innovation.

Data virtualization and data preparation are the next most common focus areas for investment, and in combination, they have the potential to improve the efficiency of self-service analytics. Data virtualization is a key enabler of providing access to multiple data sources, and along with self-service data preparation tools, it can help reduce the barriers to accessing and preparing data. This is rated by respondents as an equal barrier to self-service analytics as operationalizing the actual self-service analytics tools.

Reinforcing the point that DataOps is not just about new technology, however, investment in DataOps will also take the form of changing business processes, organizational structure or training/consulting to improve DataOps, with 84% of respondents expecting significant (38%) or moderate (46%) change in relation to processes, organizational structure, or training/consulting in the next 12 months.

Importance of DataOps

This investment in improved data management technology and business processes is being driven by the need to deliver higher-level business benefits, including greater security and compliance, increased business agility and faster time to market, as well as improved business insight and decision-making and also the development of new data-driven business applications. More specifically, DataOps is being adopted as a means of overcoming some key data management challenges, including data security and data privacy concerns and limitations related to accessing and preparing data, as well as overcoming data silos and the increasing spread of data across multiple locations.

More than three-quarters (80%) of respondents indicated that they are dealing with more than 100 data sources, while 92% of respondents reported that it takes more than a day to provision a new data environment in response to a request from a data or business analyst. As a result, 30% of respondents said it takes their organization more than a week to generate insights from raw data, and more than half said it takes more than three days. This is quite simply too slow to enable a business to be able to respond to changing business requirements.

The State of DataOps survey results also indicate that a significant proportion of time is being spent by staff finding and gaining access to data. A quarter of respondents said staff are spending more than 50% of their time doing so.

<10%	11-20%	21-30%	31-40%	41-50%	51-60%	61-70%	71-80%	81-90%	>90%
7%	13%	21%	22%	12%	13%	5%	5%	1%	1%
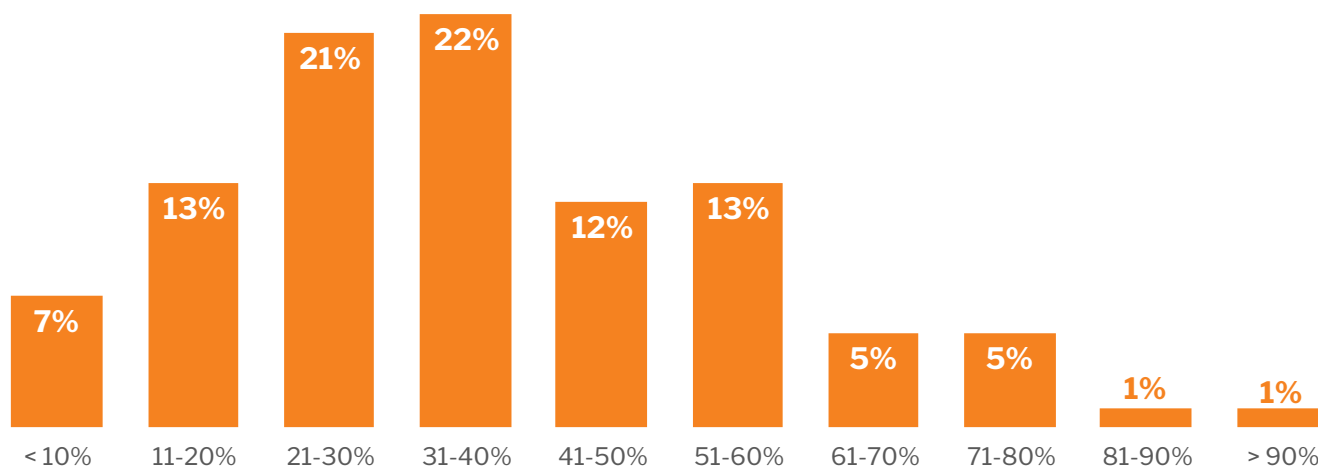
25% say staff spend more than 50% of their time finding and gaining access to data

DataOps provides some hope for overcoming these perennial data management challenges, however. No fewer than 81% of respondents agreed (32%) or agreed strongly (49%) that improved DataOps would have a positive impact on their organization's success.

DataOps Drivers

Since the term was coined, DataOps has been most closely associated with improved efficiency in the delivery of analytics projects. However, analytics is just one of the major areas that can potentially benefit from DataOps. The three major drivers for DataOps are:

- DataOps for analytics
- DataOps for operational agility
- DataOps for governance

These three drivers for DataOps are by no means independent from one another, but when asked to identify the single most significant driver for them, 53% of respondents cited operational agility benefits, 32% analytics benefits, and 14% cited governance benefits.

The results varied according to industry, however. While responses in the financial services sector were similar (54% operational agility, 30% analytics and 16% governance), those in manufacturing/construction had a greater focus on analytics (48% operational agility, 38% analytics and 12% governance), while respondents in the healthcare space had a greater focus on governance (45% operational agility, 26% analytics and 29% governance). We will drill into the specific operational, analytic and governance benefits of DataOps in the next section, but before we do so, it is worth examining the anticipated overall technological benefits, which also reflect the relative importance of the three major drivers.

Of the eight anticipated technological benefits, four are concerned with what we would consider to be operational agility benefits, with cloud migration (57%), backup/disaster recovery (52%), and development and testing (51%) rated as the three most significant potential benefits. While only one of the top eight technological benefits related to governance, compliance/risk management was rated fourth highest – by 49% of respondents – ahead of all three analytics-related benefits: artificial intelligence/machine learning (48%), BI/analytics (42%) and reporting (36%). A case could also be made for archiving (30%) to be considered a governance benefit, although we considered it to be primarily an operational efficiency concern.

Source: 451 Research survey, commissioned by Hitachi Vantara n=300
Q: Which of the following activities would benefit from improved DataOps?

Activity	%
Cloud Migration	57%
Backup/Disaster Recovery	52%
Development and Testing	51%
Compliance/Risk Management	49%
Artificial Intelligence/Machine Learning	48%
BI/Analytics	42%
Reporting	36%
Archiving	30%

■ Operational agility   ■ Analytics   ■ Governance
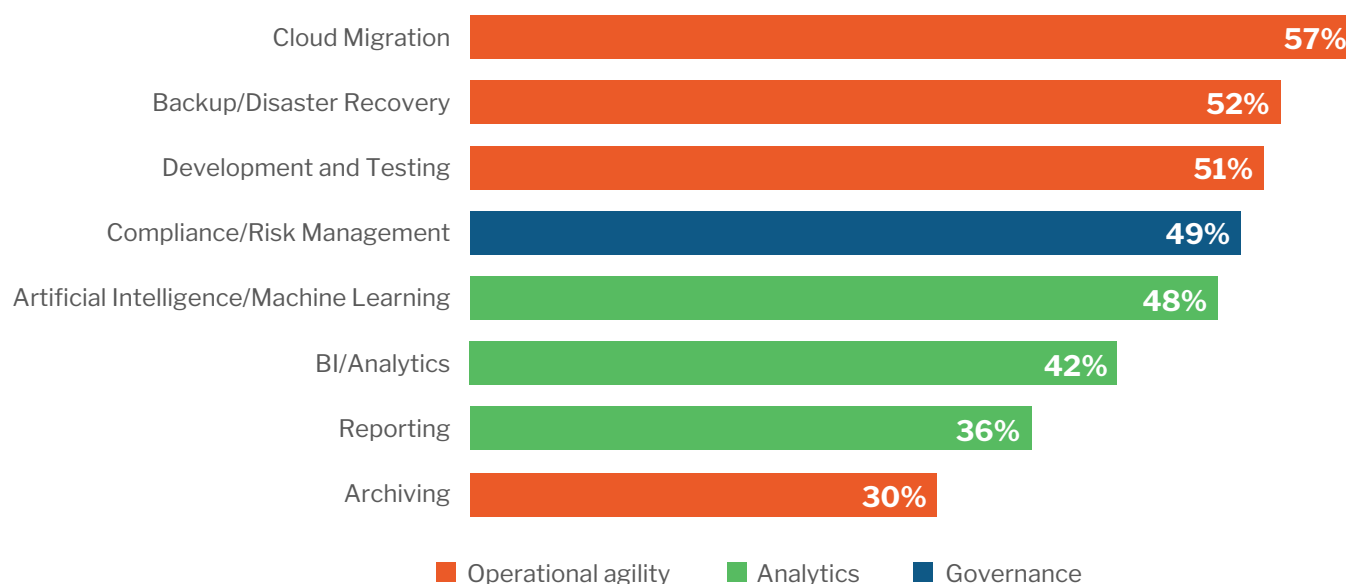
DataOps for Operational Agility

DataOps for operational agility is focused on enabling the organization to better respond to changing business requirements and improve operational efficiency. The fact that operational agility was cited as the most significant of the three primary drivers is reflective of the importance of 'Ops' in 'DataOps.' While the overall concept is clearly focused on the advantages of making better use of data, those advantages (including those that are specifically related to analytics and governance) are achieved through greater operational efficiency and agility.

When asked to specifically name the business benefits anticipated from the use of DataOps for operational agility, survey respondents cited the ability to drive efficiencies by optimizing infrastructure utilization (48%), followed by reducing data outages (36%), and freeing up data/IT staff resources through automation (16%).

The survey results also illustrate that a number of current trends are placing greater pressure on organizations in relation to achieving these operational benefits. For example, 86% of respondents stated that they have data spread across two or more cloud providers, with 55% using three or more cloud providers, 22% four or more, and 11% five or more.

Data gravity means that any significant volume of data stored in a particular environment is unlikely to be moved to another location, meaning that it may not be available for use with other analytics tools and platforms. As such, it is important to ensure that this spread of data across multiple cloud providers is managed properly. If not, fragmentation has the potential to cause analytics- and governance-related challenges because data gravity can result in data silos – isolated repositories of data that are controlled by one department and are not shared with the

rest of the organization. Data silos have the potential to lead to dark data – data that is unused for analytics purposes. Dark data, and the lack of insight into what data is being stored and processed in various locations, could lead to data risks, including compliance and regulatory challenges.

This is particularly important given that the existing investment in on-premises data lake environments is now being complemented by cloud-based data lake projects that may span multiple clouds. It comes as little surprise, therefore, that 87% of respondents rated the ability to support multiple cloud environments as critical (45%) or extremely critical (42%) to their DataOps strategy.

The challenge of data sprawl is exacerbated by the fact that an increasing volume of data is being generated on edge devices. Two-thirds of respondents said that their organization is currently generating more than 20% of its data at the edge. This figure is actually higher in the financial services (75%) and manufacturing/construction (73%) sectors, and lower in others, such as healthcare (58%).

Indeed, given the increasing volumes of data being produced by the Internet of Things (IoT), and the rate at which it is being generated, a case could be made that it is no longer a matter of enterprises deciding what data to process at the edge, but that processing data at the edge is necessary to decide what data to store and analyze in centralized infrastructure.

DataOps for Analytics

DataOps for analytics is primarily focused on enabling organizations to generate insight from their data more efficiently, through both traditional analytics and more advanced approaches including machine learning and deep learning.

The primary business benefits from DataOps for analytics anticipated by the survey respondents are greater flexibility accessing varied data sources and integrating with analytic engines (40%). This is particularly important given the trend toward the separation of compute and storage that is enabled by the use of object storage as the basis for analytics initiatives. Unlike traditional Hadoop-based deployments, the use of object storage as the basis for a data lake enables the use of multiple analytics engines. The next most cited business benefit was improving response time in relation to information requests from internal or external customers (31%), and more frequent (agile) updates to data pipelines and AI models in production (29%).

There are also analytics benefits to be had through reducing the time spent by data engineers on operational data requests, including overall faster response to business-critical data requests (70%), more time for planning and data modeling (59%), and less stress at the end of quarter/end of year or other reporting times (47%).

As an increasing number of enterprises are investing in the development of machine learning models, DataOps also provides the potential to deliver benefits specifically in relation to machine learning; 42% of respondents are currently using machine learning while another 34% are in the proof-of-concept stage. Additionally, 12% plan to use machine learning in the next year.

The barriers to using machine learning are many and varied, with accessing and preparing data topping the list (39%), followed by the ability to deploy the results into operational systems (35%), limited budget (34%), and difficulty in building and maintaining models (31%).

Among those respondents currently in deployment or proof of concept with machine learning, only 19% said that more than half of their machine learning models being developed internally are successfully being deployed into production. This is in part a matter of immaturity, but it also highlights the complexity involved in machine learning development and deployment.

The survey results also illustrate that the more mature a company is in relation to DataOps adoption, the more successful it is in terms of deploying machine learning into production, with 42% of respondents at companies where DataOps is ingrained in company culture seeing more than half of internally developed machine learning models deployed into production.

DataOps for Governance

Improving the ability to respond to regulatory requirements is one of the primary aims for DataOps for governance, alongside establishing the governance rules that can enable the acceleration of analytics initiatives. In terms of DataOps for governance, the primary business benefits anticipated by survey respondents are lower risk for compliance violations (44%), followed by faster response to legal requirements (35%), and fewer revisions due to out-of-compliance notices (21%).

The range of data governance challenges are many and varied. Training, change management, and adoption of governance technology/processes was the most popular (32%), closely followed by limited technical skills or understanding among data governance stakeholders (31%) and manual processes that slow down delivery and impact quality/consistency (28%).

Not surprisingly, those respondents who claimed to be the most mature in terms of their approach to data governance cited fewer challenges. In fact, more than a quarter of the most mature companies in relation to data governance claimed to have no data governance challenges at all, compared to 12% of all respondents; 21% with a higher DataOps maturity level also claimed to have no data governance challenges.
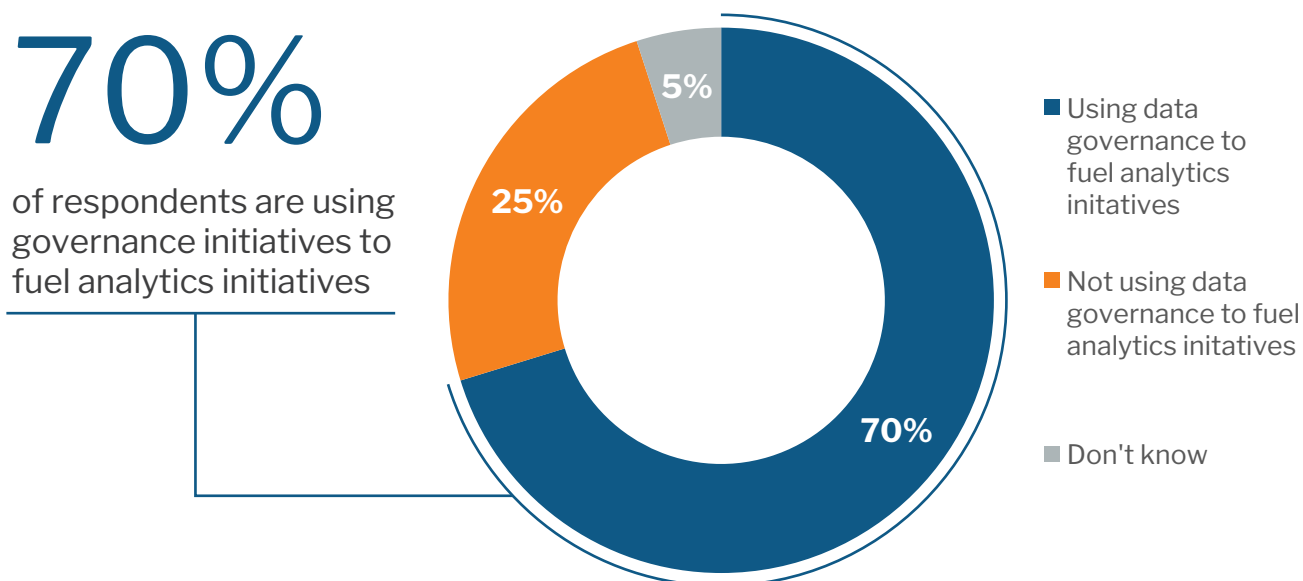
As stated above, one of the benefits of a more mature approach to data governance is acceleration of analytics initiatives. 451 Research has noted in recent years that enterprises have begun to adopt a different perspective in relation to data governance, treating it as providing the guardrails that encourage enterprises to move faster with self-service and agile analytics.

The survey results indicate that there is widespread adoption of this approach and that adoption corresponds to greater maturity in relation to data governance and DataOps. Overall, 70% of respondents stated that they are using governance initiatives to fuel their analytics initiatives. That figure rises to 82% of respondents who claimed to be among the most mature when it comes to data governance, and 81% of respondents who claimed to be among the most mature when it comes to DataOps.

70%

of respondents are using governance initiatives to fuel analytics initiatives

5%

25%

70%

- Using data governance to fuel analytics initatives
- Not using data governance to fuel analytics initatives
- Don't know

Rather than preventing access to data, this emerging attitude to data governance is rooted in providing access to the right people for the right purpose. This is reflected in the advantages cited by respondents in relation to the use of data governance for analytics initiatives. Ensuring data is distributed to appropriately authorized individuals (64%) was the most popular choice, followed by artificial intelligence/machine learning (61%) and improved overall data access (60%).

451 Take

The survey results provide evidence for the growing interest in and adoption of DataOps in recent years, and specifically the fact that enterprises are not just paying lip service to what could be considered the latest buzzword, but are also taking active steps to invest in technology, people and processes to ensure that they benefit from more agile and automated approaches to data management. Those benefits are being enabled by investment in technology, but it is also clear that the advantages of DataOps transcend the use of data management tools to also embrace cultural and organizational improvements.

While the survey results indicate that operational agility is the primary driver for most organizations, the results provide evidence that DataOps also provides advantages in relation to analytics (including machine learning) and governance. Indeed, the survey results also indicate that those three primary drivers are mutually beneficial, with operational agility advantages providing opportunities for improving analytics and governance initiatives; analytics advantages providing opportunities for improving operational agility and governance initiatives; and governance advantages providing opportunities for improving operational agility and analytics initiatives.

Survey Demographics

451 Research's DataOps survey was conducted in September and October 2019 with 300 representatives of North American organizations (with more than 1,500 employees and a minimum of 2PBs of data under management) who were directly involved in or influence the purchase, use or management of data management tools.

Nearly half (49%) of the respondents were from companies with 1,500-4,999 employees, and 51% with more than 5,000, while 38% of respondents represented companies with 2-5PB of data under management, 37% with 5-10PB, and 25% with more than 10PB.

A range of industries (financial services, healthcare, technology, manufacturing, retail, government, retail/hospitality, and communications and media) were included, as well as a range of roles (both junior and senior) covering development, testing, database management, IT operations and business intelligence. More than a quarter (30%) of respondents were CIOs, and 24% IT operations directors, while 9% were director or VP of business intelligence or business analytics.

Although we did not set out to deliberately focus on respondents that were already actively engaged in DataOps, survey respondents were mostly mature in their level of DataOps adoption: 24% declared themselves 'optimized' (with DataOps ingrained in company culture), while 39% saw themselves as 'accelerated' (with DataOps technologies and initiative scaling across multiple departments) and 28% 'emerging' (in the process of defining a DataOps strategy). Only 6% considered their level of DataOps maturity to be 'nascent' (piloting DataOps technologies and processes), while 3% admitted to being of 'low' maturity (with no DataOps strategy in place).

Source: 451 Research survey, commissioned by Hitachi Vantara. n=300
Q: How would you characterize your organization's level of maturity with respect to agile and automated approaches to data
*   management (DataOps)?*

- ■ **Optimized** –
  ingrained in company culture

- ■ **Accelerated** –
  policy has been defined

- ■ **Emerging** –
  strategy is being defined

- ■ **Nascent** –
  challenges and
  opportunities identified

- ■ **Low** –
  no strategy in place
  or in process

3%

24%

6%

28%

39%
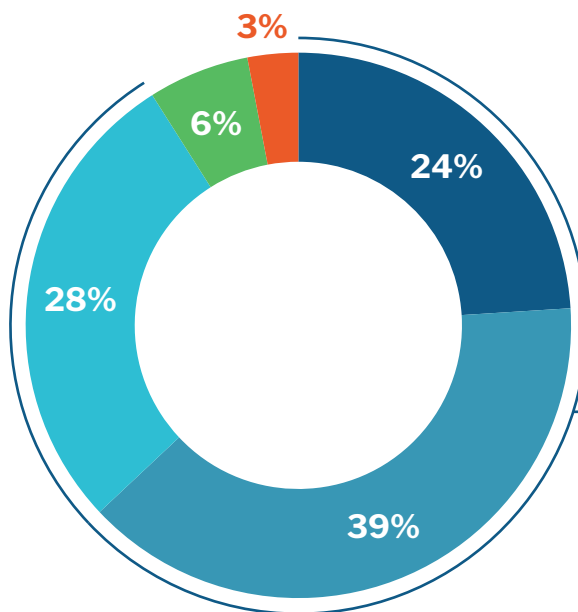
91%
of enterprises
already have, or are
in the process of
defining, a formal
DataOps strategy

About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events.

NEW YORK

Chrysler Building
405 Lexington Avenue, 9th Floor
New York, NY 10174
+1 212 505 3030

SAN FRANCISCO

505 Montgomery Street,
Suite 1052
San Francisco, CA 94111
+1 212 505 3030

LONDON

Paxton House
30, Artillery Lane
London, E1 7LS, UK
+44 (0) 203 929 5700

BOSTON

75-101 Federal Street
Boston, MA 02110
+1 617 598 7200

451 Research®