451 Research

Business Impact Brief

Agile Data Management as the Basis for the Data-Driven Enterprise

The 451 Take

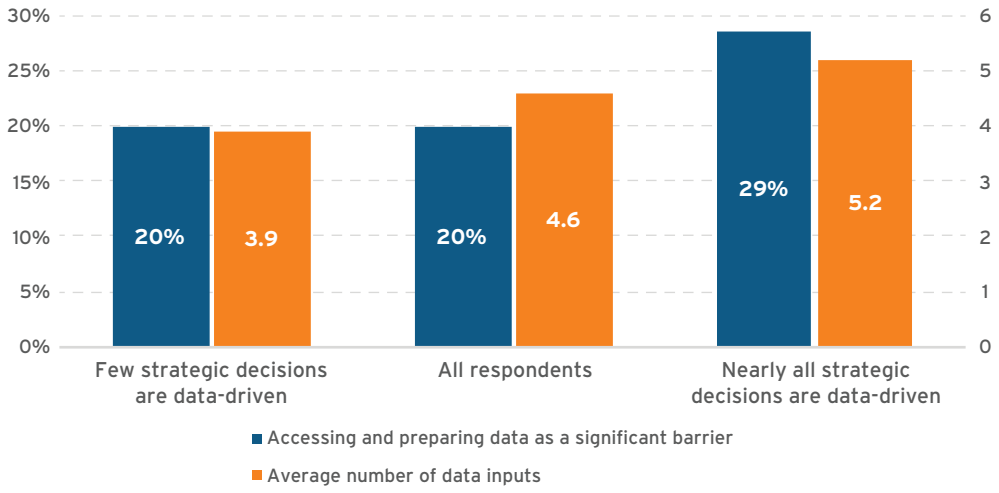
Most companies are increasing their investment in data-processing, analytics and machine-learning software with a desire to become more data-driven. Data – and the rapid processing of data – is a key driver in enabling companies to grasp the opportunities presented by digital transformation to deliver improved operational efficiency and competitive advantage.

Being data-driven requires enterprises to be more responsive to data and able to adapt to changes in data sources and formats, and in the development of new applications. While many enterprises are focused on self-service analytics and data science experiments to improve data-driven decision-making, this adaptability to change must be reflected in the entire data-processing pipeline – from data ingestion and inventory, through data blending and preparation, to data delivery, discovery and visualization.

This is easier said than done, not least because the complexity involved in integrating and managing data actually grows the more data-driven a company is. Evidence for this comes from 451 Research's recent Voice of the Enterprise: Data and Analytics survey. The results showed that while the most data-driven companies enjoy benefits such as increased focus on competitive advantage, they are also faced with more data integration and preparation overheads.

More Data Sources Means More Challenging Data Access and Preparation

Source: 451 Research, Voice of the Enterprise: Data and Analytics, 2018



On average, companies for which nearly all decisions are data-driven are dealing with 5.2 different sources of data, compared with 3.9 sources for the least data-driven and 4.6 for all respondents. More data sources means accessing and preparing data becomes more of a challenge. While 20% of all respondents cited accessing and preparing data as the most significant barrier to analytics, that figure rises to 29% among companies for which nearly all strategic decisions are data-driven.

Supporting modern data analytics projects involves creating agile data-processing pipelines that are able to rapidly and automatically integrate both structured and unstructured data from multiple sources and make it available for multiple use cases. While much attention is paid to the outputs of these data-processing pipelines – including visualizations and machine-learning models used to drive business decision-making – the importance of intelligent and automated data ingestion should not be overlooked as the first step in the journey toward enhanced business intelligence.

Business Impact Brief

Business Impact

REDUCED FRICTION. The term 'DataOps' has emerged in recent years to describe the need for more agile approaches to data management in order for enterprises to realize the potential benefits of becoming more data-driven and to reduce 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 (data management and IT professionals).

IMPROVED BUSINESS AND TECHNICAL AGILITY. Abstracting data integration logic from the execution engine enables organizations to avoid becoming overreliant on individual vendors or platforms, while also enabling greater agility in the face of industry consolidation and as environments shift between on-premises and the cloud.

REDUCED COMPLEXITY. Streamlining data operations throughout the data-processing pipeline – from data ingestion through development, training, testing and deployment – enables collaboration between the IT organization, which is typically responsible for managing data and computing resources, and the business, which retains the business domain expertise that drives analytic insight.

INTELLIGENT AUTOMATION. Metadata (literally, data about data) is growing in significance given its importance to unlocking the value of data for both analytics and regulatory compliance requirements. Specifically, metadata can be used to accelerate data ingestion in terms of understanding the data and its potential use cases, but also to automate transformations.

IMPROVED ADAPTABILITY. As enterprises increasingly adopt hybrid computing approaches, data is increasingly distributed across multiple environments, both on-premises and in the cloud, while data, business logic and computing power all potentially need to be migrated between traditional datacenter environments, public clouds, private clouds and edge devices in to order to deliver agile and adaptable analytics.

Looking Ahead

As noted above, the more data-driven a company is, the more data sources it must integrate. Data from each of these sources is likely to be delivered in different formats, meaning that it needs to be blended, transformed and cleansed before it can be used to generate business insights. Data will also be delivered via different mechanisms. Most will likely be delivered from traditional enterprise applications in batch form, but increasingly it will be generated at the edge by Internet of Things devices and delivered via stream processing, for IoT analytics.

As data is increasingly spread across multiple environments – including on-premises datacenters, in the cloud, at the edge and in hybrid environments – it will become increasingly important that the data management approach is adaptable to data being stored and processed in multiple environments. Agile approaches to data management also must be able to take advantage of multiple open-source data-processing engines, such as Apache Hadoop, Apache Spark and Apache Flink, as well as relational and nonrelational databases and other data-processing and storage technologies.

Additionally, in attempting to become more data-driven, many organizations are investing in machine-learning tools and developing ML-driven applications. The success of these projects depends on the ability of the organization to operationalize experimental data-science projects through training and testing to model deployment and management.