

Data and analytics are key to unlocking value and providing actionable insights for effective decision making in an intelligent enterprise.

## Achieving Intelligent Data Operations

June 2019

**Questions posed by:** Hitachi Vantara

**Answers by:** Stewart Bond, Research Director, Data Integration and Integrity Software, and Chandana Gopal, Research Director, Business Analytics Solutions

### Q What challenges are associated with data operations in modern data environments?

**A** Data is the lifeblood of the modern digital business. It is captured from external ecosystems of partners, customers, and devices, flowing into an organization where analytics and intelligence are used to derive insights and support decision making. These insights, when applied to business operations, in turn generate more data, and the cycle starts over again. Modern data environments have become very complex; data is no longer in a mainframe or a relational database in neatly structured rows and columns; rather, it is distributed across hybrid and multicloud environments. The increase in diversity of data types — structured, semistructured, and unstructured — demands multiple technologies to manage data at rest and data in motion, leading to multiple challenges including data integration, degradation of data quality, technical incompatibilities, and a lack of knowledge among data workers on where data is, how to access it, what it means, and when to use it, which can also make it difficult to govern data.

These challenges have generated a lot of friction within data-driven organizations, requiring multiple technologies for managing, integrating, and controlling the flow and consumption of data from the edge to the cloud and all points in between. This is causing inefficiencies because data operations are being implemented in multiple technologies and manual processes, which in turn creates another set of challenges including outdated metadata (i.e., the data about data that provides data intelligence), technical constraints, and skills gaps, resulting in an inability to leverage emerging technology and a loss of control over data. A new class of intelligent data operations platforms is emerging that can reduce friction, improve efficiencies with automation, and provide flexibility and openness with policy- and metadata-driven processes that can accommodate the diversity and distribution of data in modern environments. The challenge is to overcome the status quo and get past issues of being "too busy to change"; organizations don't have the time or money to change because the current state of data operations is costing time and money.

## Q Unstructured data can also be known as "dark data." Can organizations unlock the value of unstructured data?

**A** According to IDC's most recent Global DataSphere, which quantifies and analyzes the amount of data created, captured, and replicated in any given year across the world, unstructured content makes up 85% of new data created annually, with a compound annual growth rate of 16% for the period from 2019 to 2023. This is also a compound annual problem for organizations looking at data as an asset without knowing what data they have. For businesses to derive any meaningful insights out of this unstructured data, and gain control over what exists, structure can be created through the application of natural language processing, machine learning, and artificial intelligence (AI) and then captured in metadata. Intelligent data operations platforms can leverage this metadata to organize unstructured data and extract meaningful insights, reduce friction, and unlock the data's value.

Metadata extracted from unstructured data adds to the collective intelligence about data in an enterprise; combining metadata search and indexing capabilities with data integration capabilities allows for enrichment of unstructured data object metadata with business operations context such as customer or patient history; maintenance and repair operational records; and procurement, supply chain, and production data, opening entirely new opportunities for analytical use cases. In turn, unstructured data can add context to business operations, using metadata indexing to integrate rich content with business and field applications, such as adding product information and specifications in procurement or attaching drawings, pictures, or repair manuals to a work order for field service technicians. Intelligent data operations platforms enable the integration, access, and retrieval of unstructured data with structured data and processes from the edge to the cloud.

## Q Data lakes are as diverse as the data within them. How are organizations able to effectively manage and extract value from data lakes?

**A** Not long ago, the term data lake was synonymous with Hadoop, but this is no longer the case today. Data lakes are emerging as a diverse collection of data persistence and processing technologies that may or may not include Hadoop. Part of the movement to cloud has elevated object storage to the top of the list for data lake repositories. Where Hadoop was successful at leveraging commodity hardware to manage and process vast amounts of data by coupling compute and storage, cloud has demonstrated flexibility by separating compute from storage. This separation has allowed organizations to include object data, in-memory data, and streaming data in the scope of their data lakes, tiering data from the edge to the cloud while balancing price and performance.

Data lake operations need to have the flexibility, openness, and automation capabilities to include object stores.

Data lake operations need to have the flexibility, openness, and automation capabilities to include object stores, streaming, in-memory, Hadoop and Spark data storage and processing technologies. Data that's separated across these different technologies, in disparate locations, can be unified by an intelligent data operations platform through automation of metadata collection. This automation enables construction of metadata-driven pipelines for delivery of refined and integrated data sets to analytical and decision-centric use cases, in batch and near real time, to meet the demands of insights at the speed of business.

## **Q How can enterprises unlock the potential that is hidden in their complicated and distributed data assets and infrastructure and deliver value to the decision maker?**

**A** The amount and variety of data being generated continue to grow rapidly. According to IDC's Global DataSphere research, humans generated 33ZB of data in 2018 and are projected to generate 103ZB by 2023 (26% CAGR). The problem is that a vast majority of data that is generated today is lost. In fact, only about 2.5% of all data is actually analyzed. The biggest challenge to unlocking the potential that is hidden within data is that it is complicated, siloed, and distributed. To be effective, decision makers need to have access to the right data at the right time and with context. Although there has been a significant increase in the number of chief data officers in large enterprises, the primary responsibilities are tied to governance and efficiencies rather than value creation or innovation. According to IDC's *AI Software Platform Adoption Survey*, more than 60% respondents cited that the top challenge for data- or artificial intelligence–related projects is that they are too time consuming, especially getting data management right, and evolving business needs often further complicate the process.

Enterprises have to strategically invest in data-related projects, and use cases have to be identified with participation from a variety of stakeholders. Right now, there is so much hype around artificial intelligence that many enterprises are getting caught up in the rush to start AI-based projects without investing the time in understanding the business case or getting the underlying data right, which sets the project up for failure. Enterprises should consider the following:

- » Identify short-term and long-term data- and analytics-based use cases that will demonstrate business value with input from stakeholders at all levels (both business and IT)
- » Identify data-related assets that will be required for the project as well as time constraints
- » Deliver measurable value with short-term projects to build business cases for more expensive or longer projects

## Q How can the data science process be enhanced to deliver more value to enterprises?

**A** Data scientists occupy a unique space within an enterprise — they are intertwined with both IT and the line of business. Data scientists are typically hired by business to solve complicated problems, but their abilities are often misunderstood. Value creation using data science depends on the availability of the right data, support and investment in a data-driven culture, and the elimination of friction in internal processes that encourages collaboration between data scientists, business, and IT. One of the top frustrations cited by data scientists is that in many enterprises, their work is never operationalized. IDC's *Advanced Predictive Analytics Survey* (see *The Importance of Operationalizing Advanced and Predictive Analytics*, IDC #US42043416) found that less than half of organizations can claim that their data science work is sufficiently operationalized, and only 14% of organizations said that the output of data scientists is fully operationalized. This means that much of the work is underutilized and not likely to reach its audience, whether executives, managers, or staff who are making business decisions.

The data science process is a multistep effort, with 80% of the effort devoted to search and preparation rather than analysis. The data science workflow involves data acquisition, feature engineering, data preparation, tuning, model tuning, testing, training, selection, deployment, and monitoring. Many steps in the data science workflow are manual and time consuming and should be enhanced with intelligence. For example, technologies that provide automated feature engineering, model selection, or recommendations for data preparation will allow data scientists to focus more on value-added activities. Investing in a platform that can provide end-to-end support of the data science workflow, automate manual steps, and enable operationalization and monitoring will allow enterprises to truly leverage the power of data science. Such a platform can also allow developers to create applications that leverage data science outputs and make them accessible to the end consumer.

### MESSAGE FROM THE SPONSOR

#### We Are Hitachi Vantara

DataOps is data management for the AI era, seamlessly connecting data consumers with data creators to rapidly monetize all your data. We help you place the right data in the right place at the right time. We are analytics, industrial expertise, technology, and outcomes rolled into one great solution partner. We listen. We understand. We work with you. Get Your DataOps Advantage today with Hitachi Vantara.

**About the analysts:****Stewart Bond, Research Director, Data Integration and Integrity Software**

Stewart Bond is Research Director of IDC's Data Integration and Integrity Software service. Mr. Bond's core research coverage includes watching emerging trends that are shaping and changing data movement, ingestion, transformation, mastering, cleansing, and consumption in the era of digital transformation.

**Chandana Gopal, Research Director, Business Analytics Solutions**

Chandana Gopal is Research Director for IDC's Business Analytics Solutions market research and advisory practice. Ms. Gopal's core research coverage includes demand and supply trends in business intelligence, advanced and predictive analytics, and enterprise performance management markets.


**IDC Custom Solutions**

**IDC Corporate USA**  
 5 Speen Street  
 Framingham, MA 01701, USA  
 T 508.872.8200  
 F 508.935.4015  
 Twitter @IDC  
 idc-insights-community.com  
 www.idc.com

**This publication was produced by IDC Custom Solutions.** The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2019 IDC. Reproduction without written permission is completely forbidden.