IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Precisely

Today's demands for turning data (and data investments) into action in the enterprise heighten the call for data integrity tools and techniques. Trusted data and data intelligence don't happen by accident.

Putting Data Integrity into a Larger Context

October 2020

Written by: Stewart Bond, Research Director, and Lynne Schneider, Research Director

AT A GLANCE

KEY STATS

56% of business and data analysts indicate that lack of trust in the results or outputs of data analysis is a challenge.

KEY TAKEAWAYS

The output of data analytics is influencing decision making at all levels of the enterprise, and data integrity is needed to provide a firm foundation for confident actions.

Introduction

The future enterprise is intelligent, enabled by the organization's capacity to learn and informed by the organization's ability to synthesize information, delivering insights at scale using people and machines. It relies on data discovery, curation, integration, quality, security, and compliance to ensure the right data is delivered to the right resource at the right time — and to ensure the data is being used for the right reason. This is where data integrity can help organizations on the journey toward a future of intelligence.

Data is a core element in the future of intelligence, but it can also be an inhibitor to success. The current state of data cleanliness and quality, intelligence about data, and lack of context are impacting the organization's ability to integrate and synthesize information, restricting the organization's ability to learn from and use the data, and limiting the scale at which insights can be delivered for better, faster, and more confident decision making.

In an IDC survey of 310 business and data analysts conducted early in 2020, nearly half of the respondents indicated that a general lack of trust in data quality is a challenge for their organization, and 56% of respondents indicated that lack of trust in the results or outputs of data analysis is also a challenge. This lack of trust is not the source of the problem, but it is a symptom of the underlying scale, complexity, diversity, dynamic nature, and distribution of data in the digital economy.

IDC estimates that enterprises were responsible for over half of the 45 zettabytes of data that were created in 2019: 88% of that data was replicated, 60% of it was distributed, and 19% of it was created and consumed in real time. According to a June 2019 IDC survey of data integration software users, 95% of organizations are integrating up to 6 different types of data across 10 different types of data management technologies in hybrid cloud environments. Another symptom of complexity and lack of data integrity being observed is the inefficiency and diminished effectiveness of data-native workers. The long-touted 80/20 rule of time spent searching for and preparing data compared with the amount of time spent analyzing data is now closer to 85/15, according to the results of the 2019 data integration survey.

Context is important, and location awareness has become paramount in contact tracing and proximity analytics. However, most organizations have not been capturing locations with consistent formats. For example, it would be a challenge for many companies to map their supply chains in a way that allows for risk review in a location-based context.

To solve the puzzle and bring data integrity to the next level requires a layer of data intelligence that places the enterprise's data into a larger context, including knowing what data means as well as where it came from, how to use it, enriching it where it makes sense for business processes, and aligning it with relevant location(s). Structuring data resources and intelligence that can support inputs both internal and external to the enterprise can ensure consistent consumption of data. This in turn improves the organization's ability to synthesize information, increases the organization's capacity to learn, and enhances the organization's ability to deliver insights at scale for better, faster, and more confident decision making using trusted data with accuracy, consistency, and context.

Taking Data Integrity to the Next Level

The next level of data integrity will provide data-native workers in the organization with knowledge that can improve effectiveness and efficiencies. According to IDC's 2019 survey of data integration software users, organizations are more often unsuccessful in searching for, preparing, governing, and analyzing data than they are successful. The ability of data-native workers to access data from their platform of choice is impeded by the scale, distribution, diversity, and dynamic nature of data in the modern enterprise.

Data silos continue to exist and to grow as hybrid cloud distributes data between on-premises and multicloud environments. In 2019, 95% of organizations were integrating data across hybrid cloud environments, and IDC expects that number to be closing in on 100% because cloud migration has accelerated in recent months. Data silos create not only data divisions but also data context divisions. Similar data in each silo can have different meanings depending on the business context in which it was created and for which it was meant. Data integrity that includes context and intelligence about enterprise data, regardless of where it is or where it came from, can enable data-native workers and also improve their level of trust in the data.

It's important to note that the data in the enterprise will never be 100% clean, but that doesn't mean it cannot have integrity. Knowing how clean or, perhaps more importantly, how dirty data is can be used to improve decision outcomes. Integrity adds to data knowledge, and knowledge improves trust in the outcome. Software that can provide data quality scores to those making data-driven decisions furnishes an element of confidence or certainty in the decision based on the quality of information informing the decision. For example, machine learning requires a lot of data, and many organizations that have invested in machine learning and artificial intelligence have quickly realized they have data quality issues. Providing machines with data quality scores can also improve outcomes of algorithms if those scores are taken into consideration. In 2019, 75% of IDC data integration survey respondents using data quality management software had seen an improvement in data management and integrity metrics since implementation.

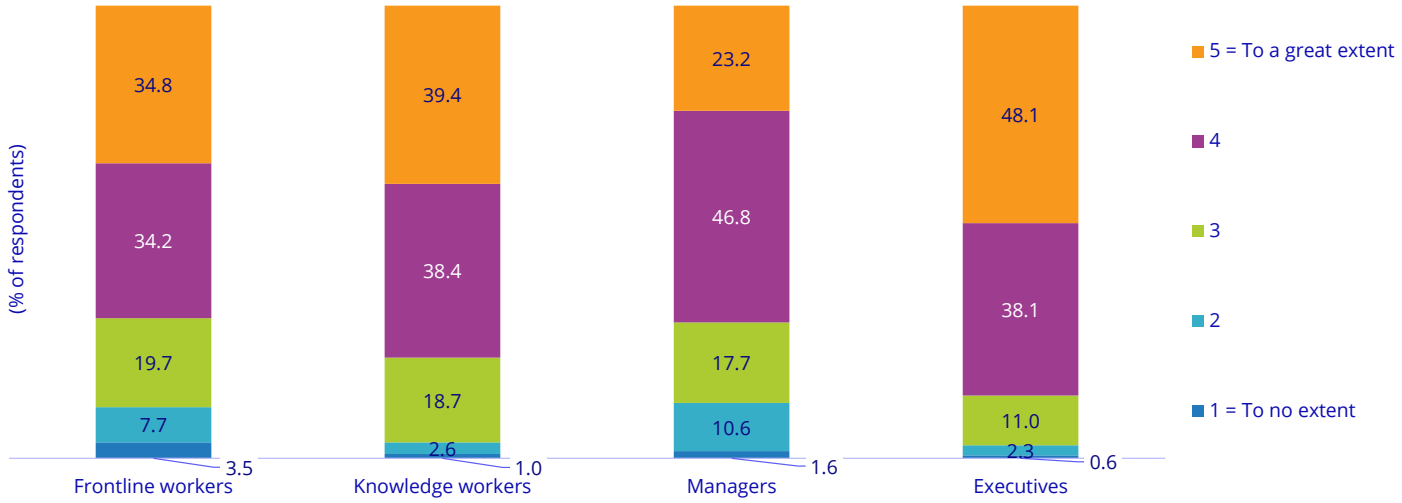
Integrity also includes how internal enterprise data meshes with external sources and indicators about those same subjects. Many enterprises choose to enrich their records with demographic, descriptive, and behavioral attributes. In addition, every asset, every transaction, every customer, and every worker will have some relationship to the physical world. Their actions and movements happen in specific places. However, most business analytics have focused on what happened and who was involved but have not considered where these events occur. Understanding the place (and point in time) helps inform the "whys" that business analysts and strategists are trying to answer. In the past, location data has been the domain of a group of specialists tasked with analyzing the physical footprint of facilities and current location of enterprise assets, among other things. Now, people in a broader variety of roles have seen the power of quality location data.

Data Integrity Trends

For an organization to be successful in this new digital economy, it must be data driven and have a good data culture, meaning it requires technology and people that can manage data capture through to consumption, with integrity. An ever-growing number of roles work with data daily to complete tasks, make decisions, and affect business outcomes. IDC has defined people in these roles as data-native workers. They exist at all levels of an organization's hierarchy, and they are pervasive in the organization in technical and business positions, from operations to strategy and from the back office to the front office. Figure 1 illustrates that the output of data analytics is affecting decision making throughout the enterprise.

FIGURE 1: ***Data Analytics Influence on Decision Making***

Q *To what extent does the output of data analytics influence or affect decision making by each of the following groups?*



n = 310

Source: IDC's Business Intelligence End-User Survey, February 2020

As organizations move toward a future of intelligence, they will need to deliver insights at scale. Scale in this instance is not only the scale of data but also the scale of data consumption across the enterprise. Data integrity is required to enable the data-native worker and improve that worker's ability to make better decisions.

Regulatory demands being placed on organizations have underscored the importance of data integrity; regulations include the ability to demonstrate compliance and protection of individuals represented in the data. Data governance is about authority and control over data. Data quality management, a key enabler of data integrity, is about the planning, implementation, and control of activities and software to ensure data assets are fit for consumption. These improvements can be used as key drivers in measuring the return on investment of data governance initiatives.

Growing Importance of Location

Location data is not just for verifying addresses and traditional GIS analytics. It is about realizing the spatial context in which interactions and transactions take place and using that context to derive greater value from data that had previously been used to plan and analyze activity. More and more enterprises are including aspects of location in their records — such as where customers live and work, as well as their mobility, or the locations of suppliers and their facilities — and increasing their intake of descriptive and demographic factors about particular locations. Location is used not only in marketing and real estate site selection; it is emerging as an important tool in areas such as contact tracing, precision agriculture, and smart emergency management. As location is incorporated into more business processes, it becomes increasingly important that the location data be precise and correct. For example, as the volume of ecommerce packages increases, the importance of shipping packages to the correct and validated customer address increases.

Global trends such as working from home, increasing volatility and unpredictability, and supply chain disruptions are casting a spotlight on how well organizations are embracing the transition to a digital enterprise while minimizing threats to business continuity. The demand for trusted data and data intelligence has increased over the past year. Data enrichment along with location context has become imperative to provide insight and visibility into day-to-day operations and for communicating effectively with employees, customers, and the entire supply chain.

Integrating and Using External Data Are Critical

The need for data integrity does not stop with internal enterprise data; it extends to any second-party and third-party data that fuels business processes and analyses. Enterprises are challenged to get integration, intelligence, and context factors right with their internal data, and 86% of senior leadership agree or strongly agree that using external data will be a critical competency over the next three years, according to IDC's 2019 *Data as a Service Survey*. These enterprises need the right tools to properly integrate internal and external data and to realize a return on the money they have invested in gathering, storing, cleansing, and purchasing data.

Considering Precisely for Data Integrity

Precisely has assembled a product portfolio unique in the data management space for its focus on data integrity. Under its prior corporate brand, Syncsort, the company focused on data quality and data integration products that provide accurate, consistent data for use in analytics and decision making. With the December 2019 acquisition of the Pitney Bowes software and data business, Precisely expanded its portfolio to include location intelligence and data enrichment.

With this expanded portfolio, Precisely is focusing its strategy on providing accuracy and consistency in data, enhanced with additional context through location and enrichment to help companies achieve data integrity. The company, recognizing that data integrity is a journey for which there is no monolithic solution, offers standalone products and the Precisely Data Integrity Suite with capabilities across the four areas that encompass data integrity: data integration, data quality, location intelligence, and data enrichment.

- **Data integration:** Precisely's data integration capabilities bring to market decades of expertise integrating complex data sources, such as mainframe and IBM i data, with next-generation technology platforms. With support for a variety of integration methods, such as change data capture, ETL, and data federation, Precisely not only offers traditional data integration into on-premises or cloud data platforms but also integrates log and machine data into IT operations and analytics platforms.

- **Data quality:** The company's data quality technology standardizes, verifies, and validates data and enables users to visualize data relationships, patterns, and trends. Robust and intelligent data profiling, entity resolution, data cleansing, address verification, data stewardship, and context graphing capabilities improve the accuracy and consistency of data.
- **Location intelligence:** Precisely's location intelligence capabilities provide the essential element of context required for data integrity by organizing, geo-enriching, and visualizing location data to fuel business processes with actionable insights. Capabilities include hyper-accurate geocoding, spatial analytics, routing, and enterprise tax delivered at-scale.
- **Data enrichment:** The company's curated, up-to-date business, location, and consumer data includes more than 400 data sets and 9,000 data attributes to enrich in-house data. Data is available for addresses, boundaries, demographics, points of interest, and streets.

Challenges

Not everyone yet sees location intelligence and data enrichment as inherent parts of data integrity. In some organizations, the responsibility to source these tools currently resides in different areas/departments. Initiatives to improve data integrity can be undermined if the root causes of data quality problems are not identified and addressed. Data trust, access, security, and compliance issues can stem from multiple root causes that include people, process, technology, and the data itself. Maximum accuracy and consistency across the population of data being used in an organization may never reach 100%, and that can be okay because trusting data for decisions includes understanding limitations and impurities in the data, providing for decisions with confidence. Precisely needs to get the market to think about data integrity differently, including integration, enrichment, location, and intelligence.

Conclusion

The future of intelligence relies on data integrity to create business value from more of an enterprise's data assets. Efficiencies can be gained in managing and developing a unified data management platform capable of improving data accuracy, consistency, and context while providing the ability to process and combine operational and external data of varying types, resulting in a larger context for data integrity. This can enable data-native workers across the enterprise to make better decisions more quickly with trusted data; help break down organizational and technical barriers; maximize accuracy, consistency, and context to facilitate synthesis of information; improve the capacity to learn; and deliver insights at scale. To the extent that Precisely can address the challenges described to deliver a unified platform across its data integrity portfolio of data integration, data quality, location intelligence, and data enrichment products, the company has a significant opportunity for success.

The future of intelligence relies on data integrity to create business value from more of an enterprise's data assets.

About the Analysts

Stewart Bond, *Research Director, Data Integration and Data Intelligence Software*

Stewart Bond is Research Director of IDC's Data Integration and Intelligence 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. Having worked in the IT industry for over 25 years, from early experience in database and application development, through solution design and deployment, to strategic architectural consulting, Stewart has worked through some significant changes in the IT industry.

Lynne Schneider, *Research Director, Data as a Service*

Lynne Schneider is Research Director leading IDC's Data as a Service (DaaS) market research and advisory practice. Ms. Schneider's core research coverage includes data sourcing and delivery services from traditional and emerging data providers along with evolving data aggregation and dissemination platforms. The breadth of coverage includes services that enable an organization to externally monetize data generated as part of the organization's ongoing operations, value-added information derived from this data, and the marketplace for combining data with other solutions.

MESSAGE FROM THE SPONSOR

The Precisely Data Integrity Suite provides confidence for better, faster decisions through trusted data with maximum accuracy, consistency, and context. Built on proven technology that 12,000 global organizations rely on for their data-driven initiatives, the Precisely Data Integrity Suite delivers value for any data integrity initiative.

With high-performance data integration and robust data quality capabilities, the Precisely Data Integrity Suite brings a foundation of accuracy and consistency to data, enabling customers to deliver more value to their customers. But it's the pivotal dimension of context that gives data integrity and sets businesses apart from the competition. The Precisely Data Integrity Suite adds context through industry-leading Location Intelligence and Data Enrichment capabilities.

The suite's modular, interoperable architecture does not require a heavy platform investment, but lets you choose from differentiated capabilities that easily deploy into existing infrastructures – opening a new world of opportunities for possibilities grounded in data.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.

5 Speen Street
Framingham, MA 01701, USA

T 508.872.8200

F 508.935.4015

Twitter @IDC

idc-insights-community.com

www.idc.com