



Predictive Analytics at Scale

An IDC White Paper, Sponsored by Micro Focus/Vertica and HPE

Author: Dan Vesset



Sponsored by:
Micro Focus/Vertica and HPE

Author: Dan Vesset

June 2020

87% of the 152 United States–based CXOs said that greater enterprise intelligence is a key priority for them over the next five years.

“It is about moving from using data to analyze performance to using data in advanced decision models to impact performance.”

Predictive Analytics at Scale

SITUATION OVERVIEW

The Imperative

After decades of progress and setbacks, the deployment and use of information technology (IT) to harness the power of data is entering a new phase. The low-hanging fruit has been picked; access to periodic business performance reports or visual and interactive dashboards, installation of yet more data warehouses or data marts, or simply building out ever-larger data lakes as discrete projects are no longer enough. Analytics is no longer about only more data or only better algorithms or only faster access to data, or only interactive data visualization.

In our ongoing interactions with C-suite executives, we hear them demanding a rethinking of what it means to have greater enterprise intelligence at scale. Their requirements are for actionable decision support (or decision automation) for everyone in their organization. In a study conducted by IDC in February 2020, 87% of the 152 United States–based CXOs said that greater enterprise intelligence is a key priority for them over the next five years. This priority is taking on more urgency as the level of global uncertainty has skyrocketed.

To achieve the goal of greater enterprise intelligence, more organizations are appointing chief data officers and chief analytics officers (sometimes as a single role). These leaders are expanding their teams of data architects, data engineers, analysts, and data scientists to support ever-greater demand for actionable insight across the enterprise — not only for executives but for everyone from managers to knowledge workers to frontline workers and increasingly for “intelligent” systems that are being deployed to automate some tactical or operational decisions.

With the new focus on raising overall enterprise intelligence comes a greater focus on modeling, simulation, and optimization — some of the techniques that enable not only creation and delivery of information but also development of insights and knowledge. In the words of one executive interviewed by IDC analysts, “It is about moving from using data to analyze performance to using data in advanced decision models to impact performance.”

Today, almost 70% of enterprises run some part of their analytics workloads in the cloud.

In the first half of 2019, 25% of overall global spending on analytics software was on cloud-based solutions, having grown at a CAGR of about 50% since 2014.

Yet too many enterprises are struggling under the complexity of today's data, analytics, and AI environment, especially in the context of broader digital transformation and economic uncertainty. Too many are saddled with legacy IT architecture that perpetuates data silos, decision silos, and knowledge silos. While this complexity may be daunting, it should be viewed as an opportunity to develop or evolve into a new, more comprehensive data, analytics, and AI strategy and platform architecture. This architecture should map to the reality of today's hybrid data environment, need for rapid development of à la carte analytic applications, scalability and performance commensurate with today's large and diverse structured and unstructured data sets and data flows,; and openness to rapidly changing analytics, AI/machine learning (ML), data integration, and data intelligence tools and services.

The Struggle with Change and Complexity

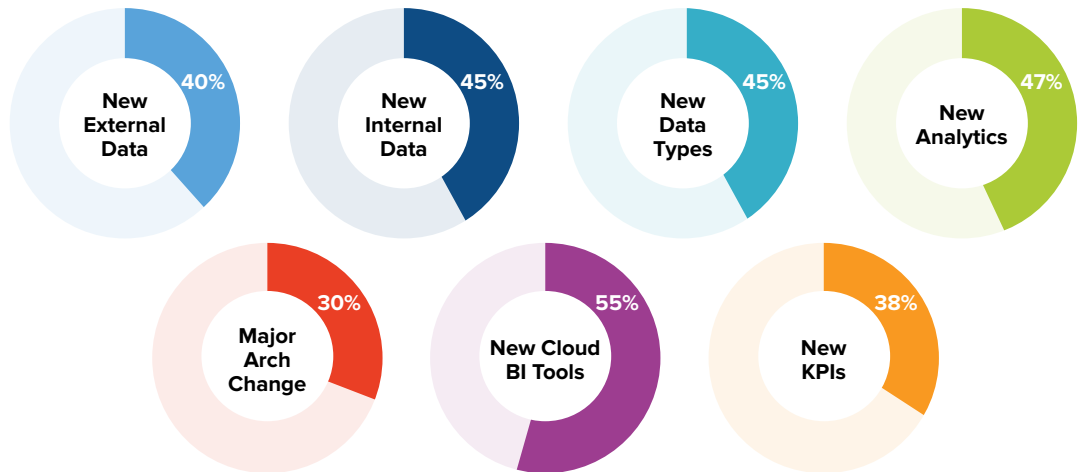
Enterprises face unprecedented and multifaceted complexity. Today, almost 70% of enterprises run some part of their analytics workloads in the cloud. In the first half of 2019, 25% of overall global spending on analytics software was on cloud-based solutions, having grown at a CAGR of about 50% since 2014. In IDC's research, we find that data engineers are, on average, working with nine unique data sources and eight unique targets per pipeline. In the United States, CXOs are telling us that their biggest challenges are a lack of necessary technology followed by the lack of appropriate analytics skills. A third of these CXOs complain of siloed data, while data professionals cite dealing with too much data as their biggest challenge. But it's not just the volume, variety, and velocity of data that affect complexity. There are also ongoing changes.

Figure 1 shows some of the changes typical enterprises are experiencing today. These include:

- » Shining a light on previous dark or dormant internal data and procuring more external data (or subscribing to data-as-a-service offerings)
- » Using new data types such as image or video or spatial data
- » Extending the use of descriptive analytics by incorporating more diagnostic, predictive, and prescriptive analytics (many based on machine learning or other forms of AI)

FIGURE 1

What Data and Analytics Changes Are Affecting Enterprises?



n = 310

Source: IDC's Business Intelligence and Analytics Survey, February 2020

30% of enterprise cite undergoing significant change to their data, analytics, and AI architecture in the past 12–18 months.

As a result, 30% of enterprise cite undergoing significant change to their data, analytics, and AI architecture in the past 12–18 months.

But the changes are not only about technology, data, and analytics. 2019 saw also a record level of CEO turnover. Our research shows that enterprises that hired a new top executive in the past three years were guaranteed to experience significant business transformation. These enterprises were also more likely to introduce new KPIs — indicative of the leadership questioning the status quo and digging into available data to uncover information that inevitably led to new questions.

In this environment, many enterprises are challenged to ensure that their data and analytics technology and processes can separate the signal from all the noise. A Nobel Laureate in economics, Herbert Simon, once said, “In an information-rich world, the wealth of information creates a poverty of attention.” Consider that he said this in 1971 and consider the transformational changes of digitization in the world that ensued in the next half a century.

62% of enterprises cite having moderate to very frequent performance issues with their analytics technology.

Challenges

As a result of the pace of change, your enterprise may be among others who cite ongoing challenges such as:

- » **Productivity:** On average, data professionals spend about 57% of their time finding, integrating, and getting data ready for analysis; 28% on analysis; and 14% on communicating the results of analysis to others.
- » **Lack of skills and technology to deploy predictive models (including those based on AI/ML) at scale into production:** Many enterprises continue to struggle with combining DataOps and ModelOps into a seamless set of processes and practices.
- » **Technical compromises:** Some enterprises have selected technology that fails to perform under the weight of the number of concurrent users, query complexity, or data volume. 62% of enterprises cite having moderate to very frequent performance issues with their analytics technology. Other enterprises have boxed themselves into suboptimal data management and analytics technology choices. For example, we have heard from many data professionals who selected a Hadoop-based data lake in the belief that this open source technology would support a much broader set of analytic workloads and use cases than it was originally intended for. They now find themselves with unexpected spending on supporting the related infrastructure, open source code base, and all the custom-coded connections from data sources to end-user business intelligence (BI), AI, and analytics tools.

Many enterprises continue to use Hadoop-based data lakes for semistructured data as a source of data for data scientists' ad hoc analysis; however, they have come to realize that data warehouses and marts based on relational technology are not replaced by the former technology because they are optimal for structured and relatively well-defined analysis and performance management use cases.

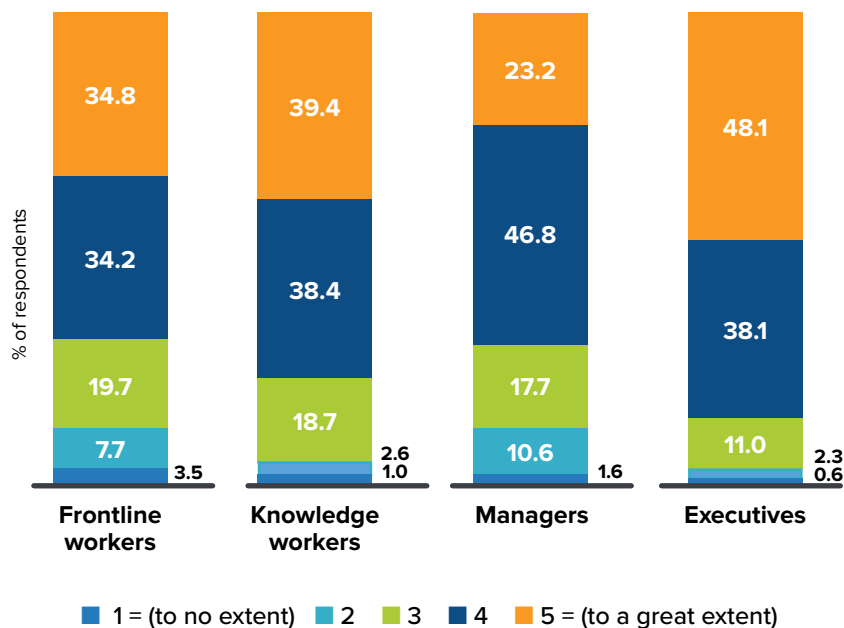
- » **Delivering insights at scale:** While some enterprises are great at provisioning analytic tools and platforms for dedicated analysts and data scientists, they fall short in ensuring the results of analysis are broadly available to the rest of the organization. In Figure 2, we show results from an early 2020 survey asking participants to respond to the question: To what extent does the output of business intelligence and analytics (BIA) influence or affect decision making by each of the following groups: frontline workers, knowledge workers, managers, and executives?

While about half of the respondents cite that executives' decisions are influenced by analytics to a great extent, this percentage drops to just over one-third of frontline workers. We are not suggesting that frontline workers should be analysts. However, their decisions in the field — related to customers or equipment or operations or finances — should also be influenced by analytics to a much greater extent. To achieve this, enterprises should develop plans to operationalize delivery of insights (developed by data scientists or business analysts) via tools and applications used by frontline employees. This can take the form of recommendations delivered at decision time, embedded into enterprise applications that run on the cloud, on premises, and at the edge.

FIGURE 2

Influence of Analytics on Actions

Q. To what extent does the output of BIA influence or affect decision making by each of the following groups?



n = 310

Source: IDC's Business Intelligence and Analytics Survey, February 2020

All these changes and resulting challenges may seem overwhelming. However, they all present an opportunity to rethink the data, analytics, and AI architecture and to redefine what it means to differentiate based on greater enterprise intelligence.

Requirements and Opportunities

All these changes and resulting challenges may seem overwhelming. However, they all present an opportunity to rethink the data, analytics, and AI architecture and to redefine what it means to differentiate based on greater enterprise intelligence. This has become evident from the actions of a select group of CIOs like the one of a large telecommunications company who, in response to IDC's questions about his company's multiyear digital transformation effort, said, "My business executives have stepped onto the path of digital transformation. As an IT leader, I have the option to continue to support their information needs as individual requests arrive, or I can be proactive in transforming our enterprise's data management and analytics architecture and solutions into an agile, scalable, and extensible platform."

This CIO articulated the challenge and the opportunity for many IT leaders tasked with prioritizing investments to address business users' ever-growing appetite for data and analytics. Many IT leaders express frustration with the difficulty of keeping pace with ever-changing business intelligence and analytics use cases from across the business.

Part of this frustration stems from a constant catchup mode that strains the business-IT relationship. For years, the typical approach to addressing internal users' needs for access to analytics assets (e.g., reports, dashboards, data warehouses, and data lakes) has been based on an attempt by IT to address one-off requests, which are often framed as a need to access a particular data set. Often these requests come with little context for the decisions business users are attempting to make using this data.

As a reaction to this need to constantly address data access "emergencies," some in IT have tried to define all the potential use cases of their business counterparts. However, this approach has also proven impractical because of the variety of potential use cases for business intelligence and analytics. Anytime, anyone, anywhere in the enterprise makes a decision, there is a use case for analytics solutions to support or augment the person making the decision or to fully automate the decision-making process. Instead of trying to identify and define each business use case, IT should focus on identifying decision-making patterns — that is, categories of decision types.

Identifying and defining these decision-making patterns enable IT to develop a technology platform, with appropriate data management and analytics capabilities for rapidly responding to any end-user request for data and analytics, including à la carte analytic applications customized for the unique needs of the organization. In fact, the previously mentioned telecommunications company developed just such an architecture and platform that now allow it to respond to any end-user analytics request within two weeks.

Several IT leaders are already embracing an approach to data management and analytics architecture that allows them to anticipate internal users' analytics needs. These enterprises have done so by rethinking how they:

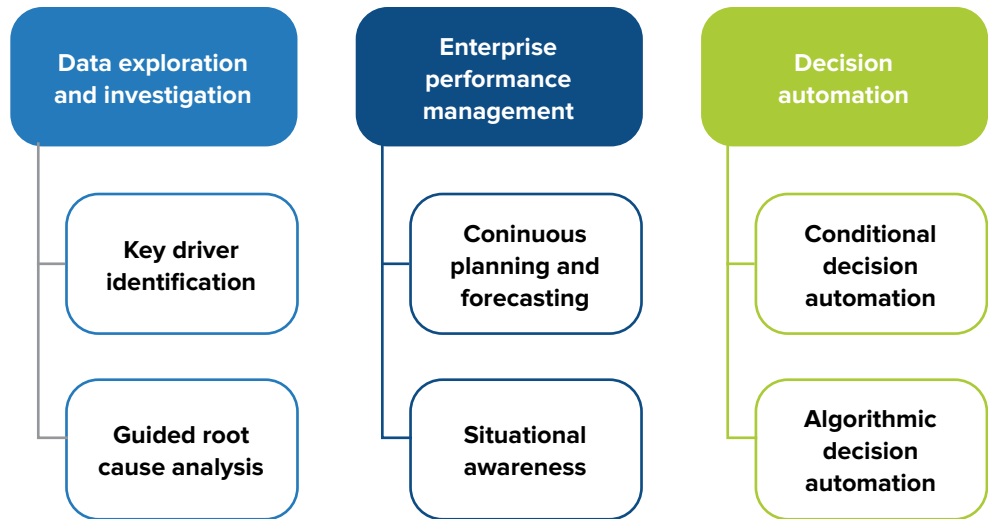
- » Approach business users' decision-making requirements rather than only data needs
- » Engage with internal stakeholders during requirements gathering dialogs

Categorizing Decision-Support Needs

These enterprises create a taxonomy of decision-making patterns, which can start with the template from IDC shown in Figure 3, which includes three categories and six subcategories of decision-making patterns.

FIGURE 3

Decision-Making Usage Patterns



Source: IDC, 2019

Data Exploration and Investigation

This decision-making pattern is about helping users understand and explain what happened in the enterprise over a given time and why it happened. Business analysts or data scientists perform this analysis using largely descriptive, diagnostic, and predictive analytics but ultimately support decision making by all levels of supervisory and managerial staff responsible for making operational decisions. The

Enterprise performance management (EPM) supports the ongoing measurement of the activities of the enterprise and of the external factors affecting it.

two primary subcategories of data exploration and investigation are:

- » **Key driver identification:** This usage pattern involves analysts (or automated systems) exploring data to identify drivers with causal effects on outputs.
- » **Guided root cause analysis:** This usage pattern is related to key driver identification. However, understanding drivers with greatest causal impact on output variables is often not enough to understand the root cause of a problem in the full context of all the internal and external factors. Today's software allows analysts' workflow to be guided by system-generated recommendations based on historical operational and behavioral data to arrive at the ultimate root cause of an issue.

Enterprise Performance Management

Enterprise performance management (EPM) supports the ongoing measurement of the activities of the enterprise and of the external factors affecting it. It provides managers and executives with better situational awareness of the current condition of the enterprise and the ability to plan and prepare in an environment of uncertainty. The two primary subcategories of EPM are:

- » **Continuous planning and forecasting:** This usage pattern involves both domain-specific and cross-domain planning and forecasting conducted on an ongoing basis to enable agile scenario evaluation and forecasting based on appropriate algorithms.
- » **Situational awareness:** This usage pattern involves instant access to or notification of the current state of the enterprise based on real-time internal and external data contextualized by historical data patterns and human expertise. Latest technology to support this usage pattern attempts to overcome the challenge of siloed, incomplete, and tardy information.

Decision Automation

Decision automation represents tactical decision making in the flow of operations. The automation, whether conditional (rules based) or algorithmic (ML based), can involve straight-through processing without any human involvement during the whole end-to-end process, or it can include the augmentation of people with lower-level task or activity automation. The two primary subcategories of decision automation are:

- » **Conditional decision automation:** The goal of this usage pattern is to receive, process, and evaluate new data continuously as it arrives, to respond rapidly to problems and opportunities, and to use optimization to make automated decisions

The three decision-making patterns are loosely aligned with different enterprise personas such as analysts/data scientists, executives/managers, and frontline employees.

about next actions. This type of decision automation provides rapid identification and response for well-known and slow-to-change conditions across a variety of processes and can be used in runtime systems for compliance.

- » **Algorithmic decision automation:** The goal of this usage pattern is to use AI algorithms and real-time data to automatically detect anomalies and opportunities, predict whether further action is needed, and apply optimization to automate or augment decision making. This type of decision automation provides the business the benefit of rapidly predicting upcoming problems or immediate opportunities where conditions change continuously and data is highly variable.

The three decision-making patterns are loosely aligned with different enterprise personas such as analysts/data scientists, executives/managers, and frontline employees. However, this persona-based alignment is not perfect. We encourage IT groups to focus on the behavior of the decision maker — that is, the decision-making patterns rather than his/her role or title.

Personas do determine data access rights and security considerations, but they shouldn't determine decision-making processes or the technologies needed to support them. For example, enterprises should not fall into the trap of considering AI technology or predictive analytics as functionality needed only by data scientists or planning capabilities as only relevant to those with the title of planner or a C-level executive.

Decision-Making Characteristics

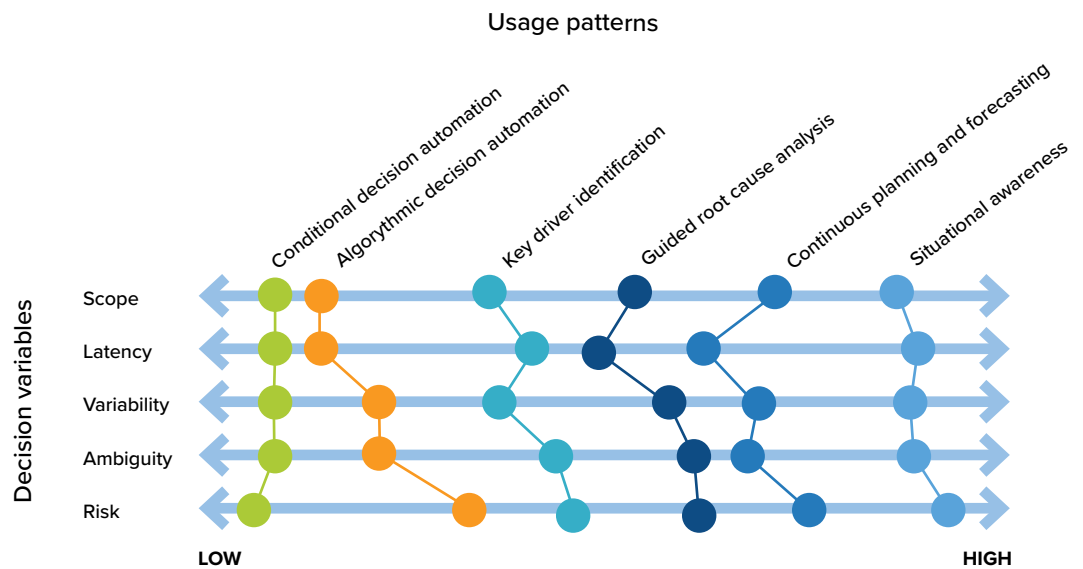
We recommend asking end users about the five decision-making characteristics as shown in Figure 4:

- » **Scope:** This characteristic defines the breadth of the impact of a given decision. Does it impact a single customer or many or all customers, or a single activity or one whole process or multiple processes?
- » **Latency:** What is the time window or time interval within which a decision needs to be made or an issue needs to be resolved? Some decisions need to be made in subseconds, while others may require weeks or months of lead time. The former is an example of real-time recommendations, while the latter is an example of a decision to acquire another company or enter a new market.
- » **Variability:** To what extent is the issue predefined versus ad hoc? Is this a regularly or consistently reoccurring decision or one that needs to be made rarely?

- » **Ambiguity:** How open ended is the issue at hand? How open to interpretation is data needed to make the decision?
- » **Risk:** What is the monetary value at risk of the decision? Decisions with narrower scope tend to have lower level of risk; however, there is not a perfect correlation between risk and scope. For example, a planning process could affect a narrow part of the enterprise but have high risk associated with compliance. Similarly, a narrowly defined tactical decision could have high reputational risk.

FIGURE 4

Decision-Making Characteristics



Source: IDC, 2019

This assessment will determine the technical requirements for the analytics architecture and platform.

Technical Requirements

The modern data, analytics, and AI architecture requires a cloud-native, services-centric approach that recognizes the need for a range of data processing engines depending on use cases. There are several “must-have” capabilities of such a platform:

- » Minimization of data movement (Whenever possible, such a technology solution must minimize or eliminate the need to move data by ensuring an appropriate balance of distributed [at the edge] and centralized [in the cloud and on-premises datacenter] data, analytics, and AI processing resources.)
- » “Out of the box” or prebuilt support for commonly used analytics, including support for AI/ML algorithms
- » Ability to extend analytic capabilities with customized and unique algorithms using the data scientists’ preferred languages and tools
- » Availability of cloud storage APIs (e.g., AWS S3 and S3-Compatible Storage)
- » Support for and integration between relational data warehousing and non-relational analytic data management, including open source Hadoop and commercial Hadoop distributions
- » Support for standard development languages and skills (e.g., SQL, Java, C++, Python, and R)
- » Support for real-time service-level agreements
- » Separation of compute and storage to enable flexibility in matching technology resources and costs to variability in analytic workloads
- » Support for Big Data processing requirements, including terabytes per second ingest/egest rate and exabyte storage capacity

Considering Micro Focus and HPE Solutions

Hewlett Packard Enterprise (HPE) and Micro Focus provide one such solution “package.”

The architecture to address the requirements laid out in this white paper must encompass optimized software and infrastructure. Hewlett Packard Enterprise (HPE) and Micro Focus provide one such solution “package.” These two technology vendors have enjoyed a 30-year partnership that has resulted in a joint analytics platform that centers around Micro Focus’ Vertica, its highly scalable columnar relational analytic database and data warehouse, deployed on HPE infrastructure on the cloud, in the on-premises core, and at the edge.

However, the partnership between the two IT companies extends beyond Vertica to include Micro Focus’ IDOL for unstructured data analysis and the company’s other software solutions in DevOps, hybrid cloud management, and security.

Like all companies in the analytics and data technology markets, Micro Focus and HPE face competition.

The solutions built on the combined Micro Focus and HPE analytics platform are frequently deployed to support AI/ML and IoT use cases and to support a broad range of other complex questions across data sources and data types. Some of these are network optimization, clickstream analytics, and route optimization, as well as smart healthcare, smart buildings, and smart agriculture industry use cases. Besides the extreme performance and scalability requirements of such systems, the joint solution offers enterprise-grade security and manageability, which are, in turn, also powered by analytics.

In addition to existing deep technical integration, experts from each organization jointly provide implementation expertise and ongoing support, allowing clients to benefit from the partnership.

As a testament to the trust that HPE and Micro Focus have placed in each other, both companies use each other's technology internally. For example, HPE Research and Development labs use Vertica for some of their most demanding data preparation, analytics, and AI model development and deployment workloads, as well as for building à la carte solutions for the largest, most demanding clients across commercial and public sectors.

Vendor Selection Considerations

Like all companies in the analytics and data technology markets, Micro Focus and HPE face competition. As always, IDC recommends all clients to go through a thorough technology evaluation process that may include third-party references and/or proof of concepts. Of special consideration should be evaluation of integration points of the joint solution with external data sources and downstream analytic tools and applications. In addition, enterprises should evaluate the fit of the Micro Focus and HPE analytic solutions based on use case patterns described previously in this white paper.

Recommendations

Even before the COVID-19 pandemic began to sweep across the world, IDC observed a greater executive-level commitment to raising enterprise intelligence. As the uncertainty has skyrocketed, the focus on improving enterprise intelligence has taken on even more urgency as enterprises seek to improve agility in planning and forecasting, optimization of operations, visibility into real-time events, and insight into addressing a new reality of human resources management. IDC's guidance is to:

- » Rethink what it means to have enterprise intelligence. It can no longer be simply about the production of reports to be delivered to a few high-level decision makers. Enterprise intelligence must be viewed as a foundational element of the enterprise culture.
- » Develop a long-term data and analytics strategy that considers various decision-making patterns.
- » Consider IT partners that provide a modern data, analytics, and AI platform that is extensible and leverages a broad partner ecosystem as no single vendor can do it all. This criterion will lead you to solutions that combine the best of open source and commercial technology.
- » Don't expect a single technology to address all requirements. One size does not fit all. SQL-based columnar MPP analytic databases have a role, as do Hadoop-based non-relational data repositories, streaming data processing tools, and a range of upstream and downstream data integration and business intelligence tools.
- » Selecting appropriate data and analytics technology is not just about finding solutions with the most compute power or storage capacity (and flexibility); consider also security, support from the solution provider, and overall total cost of ownership.
- » The TCO consideration should include how to leverage existing skills while not missing out on latest ML techniques and extend the data and analytics platform with open source components and specialized skills of data scientists.
- » Look for technology partners that have an agile strategy and technology platform that will enable your organization to make and reassess decisions about deployment options matched to your organization's wide range of decision-support and decision-automation requirements.

IDC Global Headquarters

5 Speen Street
 Framingham, MA 01701
 USA
 508.872.8200
 Twitter: @IDC
idc-insights-community.com
www.idc.com

Copyright Notice

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 2018 IDC. Reproduction without written permission is completely forbidden.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.