

## WHITE PAPER

---

# Big Data: Trends, Strategies, and SAP Technology

---

Sponsored by: SAP

---

Carl W. Olofson

Dan Vesset

August 2012

## THE DAWN OF THE INTELLIGENT ECONOMY

The intelligent economy has arrived. The convergence of intelligent devices, social networking, pervasive broadband communications, and analytics is redefining relationships among producers, distributors, and consumers of goods and services. The growth in volume, variety, and velocity of data has created new challenges and opportunities.

The information access, analysis, and management challenges of the intelligent economy can overwhelm organizations unprepared for the emerging changes. In this environment, it is not only access to data but the ability to analyze and act upon it that creates competitive advantage in commercial transactions, enables sustainable and secure management of communities, and promotes appropriate distribution of social, healthcare, and educational services.

It is not only access to data but the ability to analyze and act upon it that creates competitive advantage.

---

### In This White Paper

This IDC white paper discusses the emerging technologies of the Big Data movement. It breaks out these technologies according to their most effective roles and use cases. It also discusses why Big Data has become so important at this time and how Big Data can help enterprises reach their business goals. It considers the challenges created by Big Data and how they can be met. It identifies the relevant technology offered by SAP and shows how those technologies can help enterprises effectively leverage Big Data to their advantage.

---

### Drivers of Big Data: Why Now? What's New?

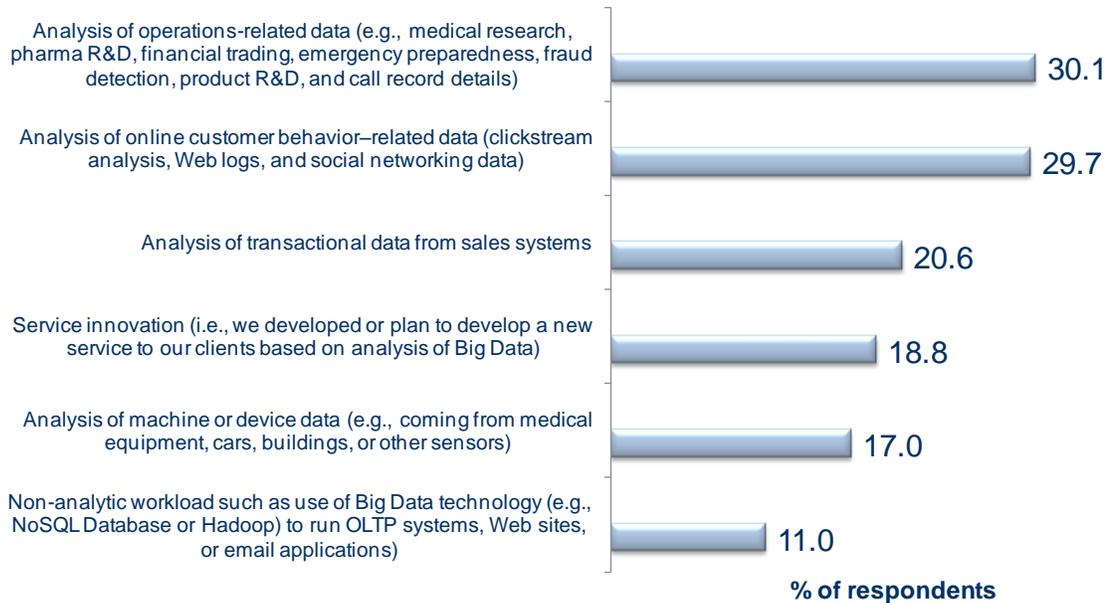
Commercial and public sector organizations are driven to invest in Big Data solutions for various reasons that span all business processes. Although data generated by social media sites garners the most attention at Big Data conferences and in press articles about Big Data, it is the analysis of operational data that is the biggest driver for using Big Data solutions (see Figure 1).

However, putting aside the rank order of the results shown in Figure 1, it is important to recognize that Big Data encompasses a broad domain of business processes, technologies, and types of expertise. As such, Big Data presents almost unlimited opportunities, but with this broad scope also come ambiguity and confusion.

**FIGURE 1**

**Top Drivers for Using Business Intelligence, Analytics, or Big Data Technologies**

Q. *What are your organization's drivers for using BI, analytics, or Big Data technologies?*



Unweighted n = 4,177

Source: IDC's 2012 Vertical IT and Communications Survey

Ultimately, Big Data solutions exist to improve decision making and to provide greater insights, faster, to decision makers at all levels of the organization.

While some equate the use of Big Data solutions exclusively with the analysis of large amounts of data by data scientists or quantitative analysts, deployments of the various Big Data solutions suggest that opportunities exist across the spectrum of various decision types and for various decision makers.

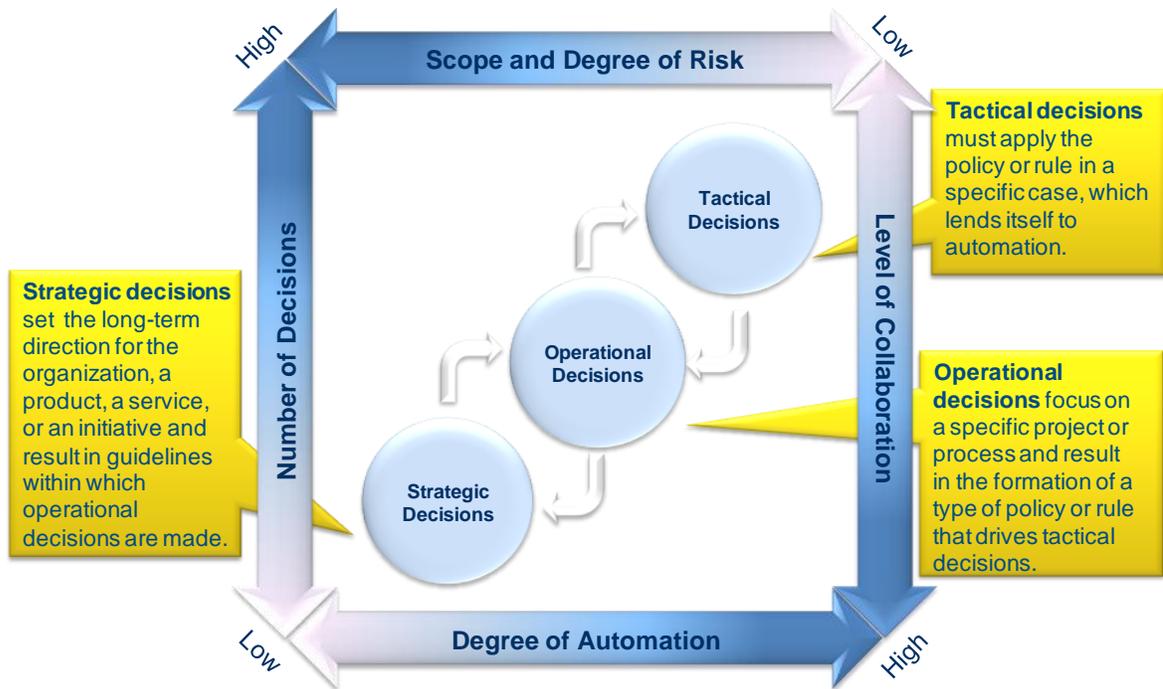
***IDC's Decision Management Framework***

One tool to evaluate these opportunities is IDC's Decision Management Framework. This framework can be applied to Big Data use cases and depicts three decision types and four primary variables of each decision type as shown on Figure 2's axes.

Strategic decisions have the highest scope and risk because of their long-term horizon and many unknown factors. There are relatively few strategic decisions; they require a high level of collaboration among internal and external decision makers, and they allow for relatively little automation. At the other end of spectrum are tactical decisions made by either frontline employees or systems. There are a lot of these decisions within a given time period, and each decision carries with it relatively little risk and is amenable to automation. These decisions are made in the field, in the flow of action, and thus there is little collaboration happening during the decision-making process. Operational decisions lie in between these two endpoints of the IDC decision management framework.

**FIGURE 2**

IDC's Decision Management Framework



Source: IDC, 2012

Each decision type is also associated with different user groups. Operational decisions are made by business or quantitative analysts and managers, strategic decisions are made by executives, and tactical decisions are made by frontline employees or automated systems, applications, or machines. The outputs of decisions of a given level become inputs for the next level of decisions. Understanding your organizational decision-making requirements is an important step toward creating a business analytics strategy that accounts for all relevant technology considerations in addition to staffing, financing, and business process considerations.

Finally, different data and technology for collecting, monitoring, managing, analyzing, and disseminating data may be required to support the various decision types and decision makers. Tactical decisions are often based on monitoring of real-time streaming data and taking action based on predetermined rules. Operational decisions may require in-depth analysis of very large amounts of multistructured data. Strategic decisions may require rapid evaluation of scenarios where instant system response enables improved risk management.

Addressing the requirements of all decision makers is a monumental task and not one that can be addressed with a single technology or within a single project.

### **Challenges Created by Big Data**

Another sign that something new is happening is highlighted by the challenges facing organizations in today's intelligent economy. In an IDC survey conducted at the beginning of 2012, both IT and business groups voiced their concern about the need to reevaluate what they are measuring to support decision-making processes within their organizations. The most frequently mentioned Big Data challenge was identified as *Deciding What Data Is Relevant*. In many cases, organizations are rethinking how to analyze existing and new data sources to change or improve decision support, decision automation, and performance management processes.

The most frequently mentioned Big Data challenge was identified as *Deciding What Data Is Relevant*.

The same organizations are also highlighting the cost of technology infrastructure and a lack of appropriate analytics and IT personnel as challenges that inhibit their ability to take full advantage of opportunities presented by Big Data. The last of the top 5 challenges is lack of business support or a lack of understanding of the benefits of Big Data by business units.

These challenges highlight the lack of established best practices for many Big Data use cases. The available data presents an opportunity to collect the data, analyze the data, and take action based on decisions influenced by the analysis. However, the achievement of this goal depends on the ability of a given organization to identify new metrics to track; the organization employing staff with the right analytics, information management, and systems management skills, a culture that is driven by analytics and accepts the results of analysis as trusted input for decision making; and the availability of appropriate technology.

## **BIG DATA TECHNOLOGY REQUIREMENTS**

### **Defining Big Data**

IDC defines Big Data technologies as *a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis*. This definition encompasses hardware, software, and services that integrate, organize, manage, analyze, and present data that is characterized by "four Vs" — volume, variety, velocity, and value (discussed in the sections that follow).

#### ***Volume: Size Is Not the Only Thing That Matters***

While the word "Big" in Big Data alludes to massive volumes of data, users must understand this as a relative term. Some industries and organizations are likely to have mere gigabytes or terabytes of data as opposed to the petabytes or exabytes of data for some of the social networking organizations. Nevertheless, these seemingly smaller applications may still require the intense and complex information processing and analysis that characterize Big Data applications.

The financial services industry demonstrates this variability. When engaging in certain Big Data activities, there may be millions or billions of records to consider, but each record may only be several bytes long (such as stock ticker information). Conversely, email archives may accumulate several petabytes of data containing valuable

customer suggestions or complaints, records of projects, legal records, contracts, and proposals. The email archive usually contains the best record of pending and current business, but it needs to be sorted and mined to find out what it contains. Another good example is product design and manufacturing, where automotive and aerospace companies, for example, may evaluate hundreds or thousands of virtual prototypes to home in on the best vehicle design. The new large-scale scientific experiments that generate petabytes of mixed data every day as input into a complex simulation model are another example.

### ***Variety: The Combination of Data Sources and Formats Is What Matters***

Variety in Big Data is a critical attribute. The combination of data from a variety of data sources and in a variety of formats is a key criterion in determining whether an application can be considered Big Data.

Big Data applications typically combine data from a variety of data sources (typically both internal and external to an organization) and data of different types (structured, semistructured, and unstructured). This is an important facet of Big Data for both technical and potential impact reasons. Combining types of information is a complex technical challenge: What is the relative importance of a tweet versus a customer record? How do you combine a large number of changing patient records with published medical research and genomic data to find the best treatment for a particular patient?

An example of this may be the mashup of internal operational data from the ERP system with semistructured data from Web log files that identifies customers' online behavior, with sentiment analysis of unstructured text from customer comments. Another example is advanced weather/climate modeling that draws on 100 years of weather data with new physical models of ocean water behaviors and CO level changes, mixing in satellite data feeds to create a real-time simulation.

### ***Velocity: Speed at Which Information Arrives and Is Analyzed and Delivered***

The velocity of data moving through the systems of an organization varies from batch integration and loading of data at predetermined intervals to real-time streaming of data. The former can be seen in traditional data warehousing and is also today the primary method of processing data using Hadoop. The latter is the domain of technologies such as complex event processing (CEP), rules engines, text analytics and search, inferencing, machine learning, and event-based architectures in general.

The key to evaluating the velocity requirements of Big Data is to understand the business/organizational processes and requirements of end users. For example, for an emergency response organization or a securities trading firm, seconds (or milliseconds) can make a difference. Similarly, real-time face recognition is a requirement for airport screening of travelers to catch criminals as they enter airports. However, Web search engines — the source of MapReduce and Hadoop — must process and mine billions of queries to determine the accuracy of their algorithms or ad matching but don't need to perform that analysis in real time. In other words, the right information at the right time with the right degree of accuracy is what's needed.

The technology infrastructure for each use case differs. There is an old adage in the infrastructure community — you can solve any problem if you throw enough hardware at it. And, when one considers the large supercomputers that have been built, or the massive clusters that have been created to address specific problems, that adage holds true.

Increasingly, however, in today's world, the need for specialized hardware is not always necessary to meet high-performance demands. The combination of high-availability clustering, scale-out file systems, multi-CPU, and multicore processors means that the performance that can be delivered today leveraging commercial off-the-shelf (COTS) components is likely to be sufficient. Socialytics applications are often delivered via cloud, making consideration of the hardware unnecessary. This is important, particularly as discussed in the Value: Capital, Operational, and Business Benefits All Matter section.

### ***Value: Capital, Operational, and Business Benefits All Matter***

In the context of Big Data, value refers to both the cost of technology and the value derived from the use of Big Data. The cost variable is important because it is a key defining factor of what's new with Big Data. Large data warehouses in financial services, telecommunications, retail, R&D, and government organizations have existed for years. Real-time data management in trading, weather monitoring, or fraud detection applications has existed for years. Unstructured content analysis in the form of text mining has existed for years. High-performance computing systems for scientific research have existed for years.

What is different now is that systems that were previously affordable and available only to government agencies or to a few of the largest companies in select industries are now available to the broader market. A combination of more widely available software and decreasing hardware prices has made these technologies more affordable.

Value also refers to the benefits derived from Big Data projects. These benefits can be broadly classified as:

- ☒ **Capital cost reduction.** A reduction in software, hardware, and other infrastructure costs
- ☒ **Operational efficiency.** A reduction in labor costs due to more efficient methods for data integration, management, analysis, and delivery
- ☒ **Business process enhancements.** An increase in revenue or profit due to new or better ways of conducting business, including improvements to commercial transactions, sustainable management of communities, and appropriate distribution of social, healthcare, and educational services

Big Data does not represent a single, homogeneous, enterprise-wide requirement. Yet the prevailing perception is that Big Data use cases represent only those dealing with very large data sets using Hadoop. For example, 40% of respondents believe that Big Data refers very large volumes of data, 26% believe that it refers to a wide variety of data, 24% believe that it refers to real-time streaming data, and 10% believe it refers to high-performance computing.

Big Data does not represent a single, homogeneous, enterprise-wide requirement.

## The Big Data Technology Landscape

IDC classifies Big Data technologies in two distinct buckets: one for Big Data in motion, the other for Big Data at rest.

### ***Big Data in Motion***

Big Data in motion represents rapidly streaming, high-volume data that must be acted upon as it arrives. Examples include stock transaction data, smart meter data, RFID data into real-time inventory control systems, and so forth. In such cases, three classes of operation are associated with the data.

In Big Data in motion use cases, data is received, filtered, and regularized (put in a consistent or readable format). This is normally done by the receiving application. The system determines if some response is required. This can involve a complex event processing engine, taking the new data, applying the new data against retained data (including both cached data from the stream and data kept in a very fast [typically in-memory] database), and determining whether or not a defined event has taken place. If a defined event has taken place, the CEP engine triggers an action, which is an application response to the event.

Big Data in motion requires technology that can receive, format, and respond to the data as fast as it comes in. This involves intelligent high-speed data movement and transformation technology as well as in-memory and CEP technology.

### ***Big Data at Rest***

Most Big Data discussions revolve around the Big Data at rest class of Big Data technology, which involves the ability to collect data as fast as it comes in, order and transform the data, analyze the data, and put the data in a state in which meaningful search, mining, discovery, query, and reporting may be done against the data.

Big Data at rest includes what we call "structured" and "unstructured" data. Lately, many experts have objected to these terms, pointing out that what we call "unstructured" data really does have structure — it is simply not determined by schemas or program code. To deal with this issue, we may consider the classifications illustrated in Figure 3.

IDC classifies Big Data technologies in two distinct buckets: one for Big Data in motion, the other for Big Data at rest.

**FIGURE 3**

"Structured" Versus "Unstructured" Data

	Type	Container
"Structured"	Extrinsic Explicit Structure	<b>Fixed or Variable Schematic Database (e.g., RDBMS, OODBMS)</b> The structure is declared and managed explicitly by a schema manager apart from the data itself.
	Extrinsic Implicit Structure	<b>Non-Schematic Database (Key-Value Store)</b> The structure is known to the application program but is not explicitly managed.
"Unstructured"	Intrinsic Explicit Structure	<b>Tagged Format Data such as XML, CDF, standard exchange format files, etc.</b> The structure is present in the data in the form of headings, tags, or other established forms of labeling.
	Intrinsic Implicit Structure	<b>Content in Human Language</b> The structure is not explicitly declared but may be inferred from the data itself, the chief example of this being human language, the semantics of which are to be found in syntax and grammar.

Source: IDC, 2012

Both NoSQL and SQL databases have important roles to play in the Big Data world. NoSQL databases are especially good at supporting the "variety" dimension of Big Data, accepting data from multiple sources in multiple formats, and then permitting program code to sift through, filter, and organize the data. This is especially common in Hadoop applications. SQL databases are especially good at handling large volumes of data that have a consistent, known structure, enabling regular reporting, mining, and repeated analysis on such data.

IDC considers RDBMSs that feature dynamic scalability, can handle very large databases, and can process requests against such databases very rapidly as Big Data SQL DBMSs. These RDBMSs have evolved from older forms to acquire these characteristics.

NoSQL is another story. The term, as commonly used, is ambiguous and is applied against a wide range of DBMS types, which are illustrated in Table 1.

Each DBMS type serves a particular purpose, and often, multiple databases of different types may be present together in the same system as part of a sensible flow of Big Data operations.

**TABLE 1**

NoSQL DBMS Types

Paradigm	Usage
Key-value	Caching shared data, creating ad hoc data spaces
Grid-based key-value	Very large ad hoc scalable data collections for analysis
List processing	Collecting large amounts of unpredictably structured data for analysis
Graph	Managing unlimited levels of relationships of data entities with rapid traversal and search capability
"NewSQL"	Accepting data to which a structure is subsequently applied, or from which a structure is inferred, to support relational queries without a predefined schema

Source: IDC 2012

## Common Misconceptions About Big Data

The market for Big Data solutions has emerged relatively recently. There are many vendors in the market today that are trying to jump onto the Big Data bandwagon. This creates confusion and generates misconceptions about what Big Data is and what Big Data technologies can do. A few of the most common misconceptions are:

- ☒ **Big Data analytics is all about doing something new and different that wasn't possible before the latest technology was developed.** The idea that Big Data is just about doing something new is also false. Big Data concepts have been around for many years. What has changed is the economics by which Big Data concepts can be deployed. What has changed is that we now have the capability of having computer-assisted discovery of relationships between very large datasets that encompass a very wide variety of sources of data. What has changed is the awareness that analytics can create competitive advantage if the right information is provided to the right decision makers at the right time using the right tools.
- ☒ **Big Data technology is all about Hadoop (and, by extension, any MapReduce environment) environments, regardless of the workload or use case.** We can also aggregate under this misconception related beliefs that relational databases can't scale to very large volumes so can't be considered Big Data technologies or that the era of schematic DBMS is over and that schema development only gets in the way of Big Data deployment. Another frequent

misconception about Big Data is that one technology, such as Hadoop, can handle all Big Data processing requirements. The reality is that one must carefully choose the right technology for the job. Just as no one wrench can tighten any nut, no one Big Data technology can solve every Big Data requirement. While NoSQL databases are gaining in popularity among Big Data use cases, relational databases continue to play an important role. While Hadoop is certainly gaining a lot of mindshare in the market, it is neither the only form of data management nor the only implementation of the MapReduce functions.

- ☒ **Big Data is only about very large data volumes and, by extension, it's primarily about data.** Large data sets are certainly a key part of the Big Data market trend. In fact, 40% of organizations identify Big Data with very large amounts of data. But there are also other characteristics, such as real-time or streaming data and a wide variety of data of multiple types or formats. Some Big Data technology deployments address one of these three characteristics of Big Data, while other deployments address two or all three characteristics.
  
- ☒ **Big Data is just a new buzzword for data mining.** Data mining refers to a set of analytic techniques that can be used to analyze large data sets. Some of these techniques have been used for centuries; others are more recent. However, Big Data, as defined by IDC and most other market observers and participants, is a much broader topic that encompasses data collection, data management and organization, data analysis, information access as well as operational workloads, and use cases that use some of the same new and established Big Data technologies.
  
- ☒ **Big Data is *the* challenge.** Perhaps the most important of the current misconceptions is that deploying Big Data technology will in and of itself solve business problems, increase revenue, decrease costs, and attract customers. Storing large amounts of data, whether in a relational database or a Hadoop cluster, is not an end in itself. The *if we build it they will come* approach has never led to successful technology deployments, and it will not succeed in Big Data. Neither is analysis of data an end in itself. There are plenty of examples of great, insightful, and timely analysis by brilliant data scientists — analysis that either does not reach or is ignored by the relevant decision makers or analysis that simply misses its mark because it discounts the behavioral variables in human interaction. An example is the recently highly publicized case of a large retailer that developed a highly accurate predictive score to identify a customer segment but bungled the marketing campaign to the selected customers because it failed to anticipate the privacy concerns of the audience.

It is important to understand these misconceptions to avoid becoming caught up in unproductive arguments about the generalized virtues of one technology versus another. The reality is that for most sizable organizations, multiple Big Data technologies need to coexist and address workloads and use cases for which they were optimized.

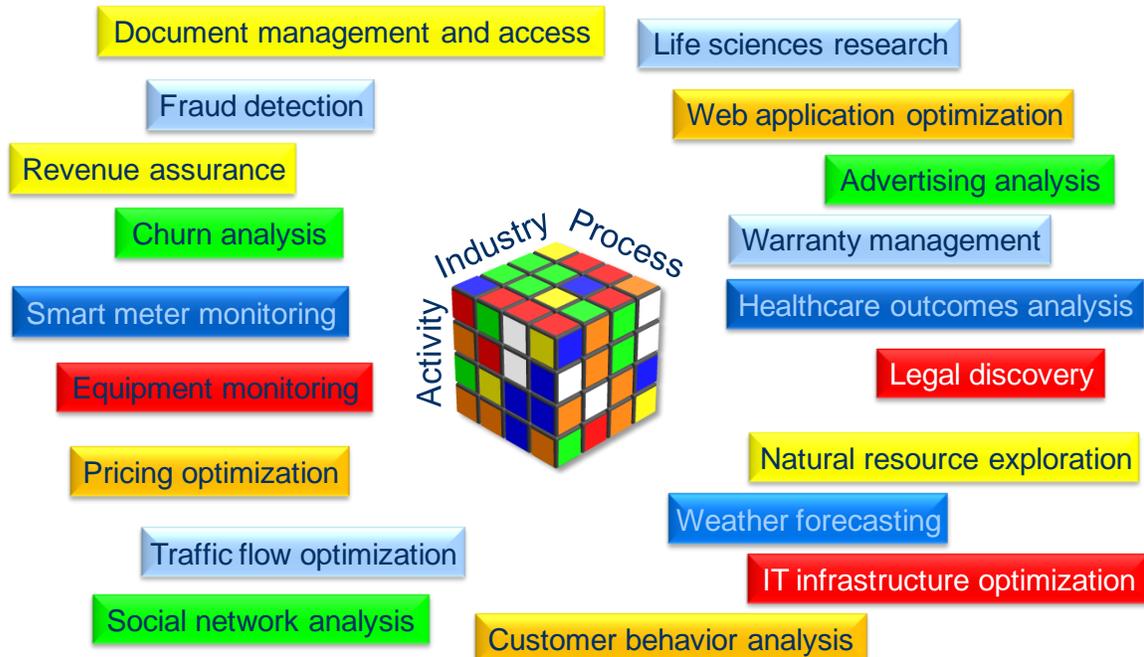
For most sizable organizations, multiple Big Data technologies need to coexist and address workloads and use cases for which they were optimized.

## Sample Big Data Use Cases

The classes of use cases of Big Data solutions vary widely. Figure 4 illustrates common Big Data use cases that we have observed in the market today.

**FIGURE 4**

### Big Data Technology Use Cases



Source: IDC, 2012

Three dimensions on which to evaluate such use cases — activity, business process, and industry — are discussed in the sections that follow.

### **Activity**

Not all applications of Big Data technologies are for analysis of data. Some are used for deploying Web sites for social media or gaming applications, and others are used for large content stores that provide information access to massive amounts of documents. Examples include:

- ☒ Analytics (e.g., data mining, multi-dimensional analysis, data visualization)
- ☒ Operations (e.g., running a Web site, processing online orders)
- ☒ Information access (e.g., search-based access to information, normalization, and access across content and data sources)

### ***Business Process***

Big Data technologies are being deployed in support of processes within commercial, nonprofit, or government organizations. The challenges and problems organizations face are not Big Data challenges but rather business or organizational challenges that are impacted by Big Data. Big Data technology deployment cases can be found across business processes such as:

- ☒ Customer relationship management (sales, marketing, customer service, etc.)
- ☒ Supply chain and operations
- ☒ Administration (focused on finance and accounting, human resources, legal, etc.)
- ☒ Research and development
- ☒ Information technology management
- ☒ Risk management

### ***Industry***

In addition to cross-industry processes such as finance, marketing, and information technology management, there are a multitude of industry-specific applications. A short sample of these includes:

- ☒ Logistics optimization in the transportation industry
- ☒ Price optimization in the retail industry
- ☒ Intellectual property management in the media and entertainment industry
- ☒ Natural resource exploration in the oil and gas industry
- ☒ Warranty management in the manufacturing industry
- ☒ Crime prevention and investigation in local law enforcement
- ☒ Predictive damage assessments in the insurance industry
- ☒ Fraud detection in the banking industry
- ☒ Patient treatment and fraud detection in the healthcare industry

The suppliers of Big Data technology solutions to address these and other Big Data use cases range from small specialty firms to large diversified technology firms with broad portfolios of related products. One of the key vendors in this market is SAP AG.

## **Big Data Technology from SAP**

SAP offers a range of technologies that address Big Data use cases and requirements. These SAP technologies span the spectrum of IDC-defined Big Data criteria such as very large 100TB+ data sets, smaller but fast-growing data sets, real-time streaming data sets, and multistructured data sets. Some of these technologies handle data at rest, while others focus on data in motion. Some of the technologies are memory based, others are disk based and utilize complex event processing, and others employ content analysis software. A sample of SAP's products that address Big Data requirements are discussed in the sections that follow.

SAP technologies span the spectrum of IDC-defined Big Data criteria such as very large 100TB+ data sets, smaller but fast-growing data sets, real-time streaming data sets, and multistructured data sets.

### ***SAP HANA***

The SAP HANA platform makes possible instant analysis of large amounts of multistructured data and the embedding of analytics into operational applications. The SAP HANA appliance enables organizations to analyze their business operations using large volumes of detailed operational data in real time while business is happening. Operational data is captured in memory and made available for instant analysis, eliminating the typical lag time between when data is captured in business applications and analysis of that data from reporting and analysis systems. This in-memory appliance combines SAP software with hardware from the company's strategic hardware partners.

### ***SAP Sybase IQ***

Focused on large amounts of data at rest, the Sybase IQ columnar and relational database technology enables integration and management of data in a single data warehouse. It is deployed for use cases that require a RDBMS that can scale dynamically while still providing very fast response times for queries. SAP Sybase IQ, with its shared everything clusters, represents such a system and has been shown in live deployments to scale to over 1PB. Use cases exist in which a SAP Sybase IQ data warehouse coexists with a Hadoop deployment where the latter is used for preprocessing of large amounts of multistructured data that is subsequently loaded into Sybase IQ for more in-depth structured analysis and long-term trending analysis.

### ***SAP Sybase Event Stream Processor***

For handling Big Data in motion, SAP offers a CEP engine and also features the ability to move data rapidly for enterprise state analysis using the SAP Sybase Replication Server. Together, these products enable a CEP-based real-time event-driven system that operates against streaming Big Data backed up by established enterprise data. SAP Sybase Event Stream Processor (ESP) is often deployed in applications requiring real-time decision making such as financial securities trading systems.

### ***SAP Enterprise Information Management***

Enterprise information management (EIM) solutions from SAP help organizations manage Big Data from a variety of data sources. In particular, SAP Data Services delivers ETL, data quality management, and text data processing capabilities for

structured or unstructured data residing in databases, data warehouses, or distributed file systems such as Hadoop. Master data management solutions, such as SAP Master Data Governance and SAP NetWeaver Master Data Management, help organizations create and govern master data across the enterprise. And SAP Information Steward enables business users to profile and monitor data quality, enabling organizations to improve the effectiveness of Big Data across operational, analytical, and governance initiatives.

### ***SAP BusinessObjects Business Intelligence***

The SAP BusinessObjects BI portfolio enables organizations to gain valuable insight from Big Data stored in databases, data warehouses, and Hadoop. In particular, the SAP BusinessObjects BI platform provides a single semantic layer for reporting, dashboards, ad hoc queries, multi-dimensional analysis, and data exploration. New solutions, such as SAP Predictive Analysis and SAP Visual Intelligence for SAP HANA, enable business users to manipulate data quickly. As a result, organizations can anticipate or to react to events faster, enabling rapid action for competitive advantage.

### ***SAP Analytic Applications***

Analytic Applications from SAP are purpose built to address industry or line of business—role specific needs to better manage everything from business planning, regulatory compliance, and profitability to the daily monitoring of KPIs. Many of the performance metrics are impacted by large volumes of structured and unstructured data generated both within the organization and outside the organization. Exploiting the capabilities of the SAP BusinessObjects BI portfolio as well as SAP HANA, analytic applications can empower executives and business employees to derive intelligent information to support decision making, analyze root causes, simulate scenarios, and make predictions. SAP Sales Analysis for Retail and SAP Smart Meter Analytics are examples of analytic applications powered by SAP HANA that deliver real-time analysis based on information sourced in high volumes from various different sources.

## **CHALLENGES AND OPPORTUNITIES**

Big Data represents a large and growing opportunity to improve existing business processes, to launch new lines of business, to change interactions with customers, and to reevaluate how and why data is being analyzed in support of a broad range of decision-making processes. However, there are also many misconceptions and challenges facing Big Data technology users and vendors.

As one of the largest IT vendors, SAP is in a strong position to capitalize on the opportunities presented by Big Data requirements. However, in its quest for success, SAP faces competition from established vendors and a new set of start-ups. The latter group has been focused on providing their niche solutions to a few industry segments such as social media, entertainment, and publishing. The broader mass market for Big Data has not yet materialized, which makes it imperative for SAP to carefully select appropriate technology that supports the early Big Data use cases of its customers across traditional and emerging industries.

As one of the largest IT vendors, SAP is in a strong position to capitalize on the opportunities presented by Big Data requirements.

## RECOMMENDATIONS

In today's intelligent economy, success in the form of better performance is increasingly defined by having the necessary people, processes, and tools to innovate, to provide customers with better products and services, to optimize operations, and in general to have the ability to act faster and with greater insight within ever-shorter decision windows in the face of uncertainty. To achieve these goals, organizations should consider, among other efforts, the following best practices of their peers or competitors:

- ☒ Develop a business analytics and related Big Data strategy that includes an evaluation of your organization's decision-making processes as well as groups and types of decision makers.
- ☒ Recognize that even within the *analyst* user group, there are a range of roles with different tasks and capabilities. Identify information and technology requirements of each user group.
- ☒ Experiment with new metrics, new KPIs, and new analytic techniques to look at both existing and new data differently and uncover opportunities for incremental improvements as well as innovation. This may require setting up a separate Big Data analytics team that is given the freedom to experiment and innovate.
- ☒ Recognize that one size (technology) does not fit all (Big Data requirements). Different workloads, data types, and user types are best served by technology purpose built for a specific use case. Hadoop may be the best choice for large-scale Web log analysis, but it certainly is not the technology of choice to analyze real-time streaming data. A columnar RDBMS is great at structured analysis of large amounts of operational data, while an in-memory analytic engine would be best suited for rapid ad hoc evaluation of alternative scenarios.
- ☒ Consider the performance impact of enabling technologies such as in-memory computing and columnar databases. Both have entered the mainstream and either separately or even more so in combination enable performance improvement of existing business analytics software or deployment of completely new software that was not feasible before.
- ☒ Evaluate both business and IT benefits. Some IT departments are keen on consolidating the number of technologies to minimize ongoing maintenance cost and range of required skills. Yet best practices suggest that it is specialization that drives differentiation and subsequent competitive advantage.
- ☒ Understand the misconceptions about Big Data to avoid becoming caught up in unproductive arguments about the generalized virtues of one technology versus another. The reality is that for most sizable organizations, multiple Big Data technologies need to coexist and address workloads and use cases for which they were optimized.
- ☒ Recognize that the goal of any Big Data project is not the collection of a large amount of data or the employment of an army of data scientists. Big Data projects must support concrete business needs, and they must provide

actionable information to decision makers, whether line-of-business executives or automated systems.

---

### **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 2012 IDC. Reproduction without written permission is completely forbidden.