IBM Global Business Services
Business Analytics and Optimization

Executive Report

In collaboration with *Saïd Business School at the University of Oxford*

Saïd Business School
UNIVERSITY OF OXFORD

IBM Institute for Business Value

Analytics: The real-world use of big data

How innovative enterprises extract value from uncertain data

IBM Institute for Business Value

IBM Global Business Services, through the IBM Institute for Business Value, develops fact-based strategic insights for senior executives around critical public and private sector issues. This executive report is based on an in-depth study by the Institute's research team. It is part of an ongoing commitment by IBM Global Business Services to provide analysis and viewpoints that help companies realize business value.
You may contact the authors or send an e-mail to *iibv@us.ibm.com* for more information. Additional studies from the IBM Institute for Business Value can be found at *ibm.com/iibv*

Saïd Business School at the University of Oxford

The Saïd Business School is one of the leading business schools in the UK. The School is establishing a new model for business education by being deeply embedded in the University of Oxford, a world-class university, and tackling some of the challenges the world is encountering. You may contact the authors or visit: *www.sbs.ox.ac.uk* for more information.

By Michael Schroeck, Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano

"Big data" – which admittedly means many things to many people – is no longer confined to the realm of technology. Today it is a business priority, given its ability to profoundly affect commerce in the globally integrated economy. In addition to providing solutions to long-standing business challenges, big data inspires new ways to transform processes, organizations, entire industries and even society itself. Yet extensive media coverage makes it hard to distinguish hype from reality – what is really happening? Our newest research finds that organizations are using big data to target customer-centric outcomes, tap into internal data and build a better information ecosystem.

The term "big data" is pervasive, and yet still the notion engenders confusion. Big data has been used to convey all sorts of concepts, including: huge quantities of data, social media analytics, next generation data management capabilities, real-time data, and much more. Whatever the label, organizations are starting to understand and explore how to process and analyze a vast array of information in new ways. In doing so, a small, but growing group of pioneers is achieving breakthrough business outcomes.

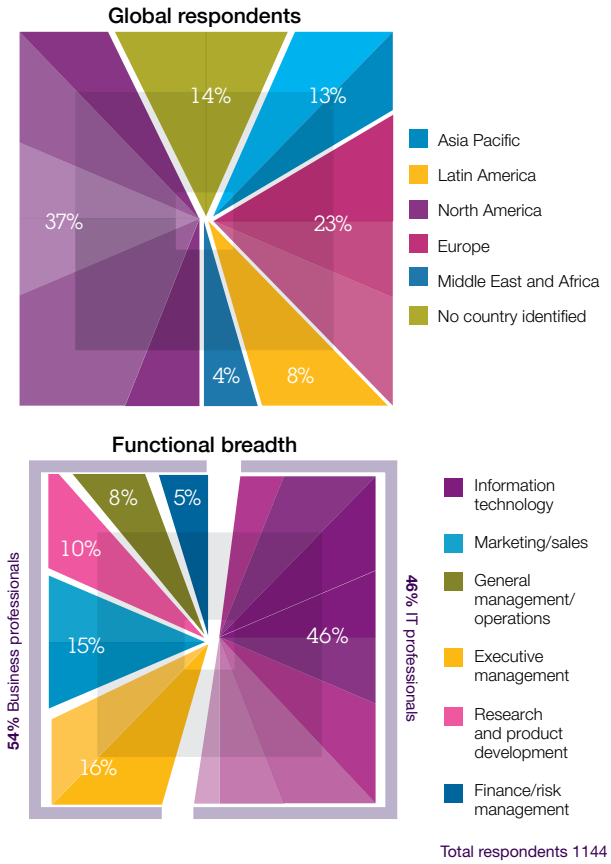
In industries throughout the world, executives recognize the need to learn more about how to exploit big data. But despite what seems like unrelenting media attention, it can be hard to find in-depth information on what organizations are really doing.

So, we sought to better understand how organizations view big data – and to what extent they are currently using it to benefit their businesses. The IBM Institute for Business Value partnered with the Saïd Business School at the University of Oxford to conduct the 2012 Big Data @ Work Study, surveying 1144 business and IT professionals in 95 countries, and interviewing more than two dozen academics, subject matter experts and business executives.

About this study

The IBM Institute for Business Value and the Saïd Business School at the University of Oxford partnered to develop this report. It is based on the Big Data @ Work Survey conducted by IBM in mid-2012 with 1144 professionals from 95 countries across 26 industries. Respondents represent a mix of disciplines, including both business professionals (54 percent of the total sample) and IT professionals (46 percent). Respondents self-selected to participate in the web-based survey.

Study findings are based on analysis of survey data, and discussions with University of Oxford academics, subject matter experts and business executives. IBM is the primary source of study recommendations.



We found that 63 percent – nearly two-thirds – of respondents report that the use of information (including big data) and analytics is creating a competitive advantage for their organizations. This compares to 37 percent of respondents in IBM's 2010 New Intelligent Enterprise Global Executive Study and Research Collaboration – a 70 percent increase in just two years.[1]

As an increasingly important segment of the broader information and analytics market, big data is having an impact. Respondents whose organizations had implemented big data pilot projects or deployments were 15 percent more likely to report a significant advantage from information (including big data) and analytics compared to those relying on traditional analytics alone.

One surprising study finding is the relatively small impact of social media data on the current big data marketplace. Given the extensive press coverage about social data's impact on customer experiences, it would be easy to believe that big data means social media data, but only 7 percent of respondents defined big data that way. And fewer than half of respondents with active big data initiatives reported collecting and analyzing social media data; instead, respondents told us they use existing internal sources of data in their current big data efforts.

So what makes today's big data activities different? Some organizations have already been handling big data for years. A global telecommunications company, for example, collects billions of detailed call records per day from 120 different systems and stores each for at least nine months. An oil exploration company analyzes terabytes of geologic data, and stock exchanges process millions of transactions per minute. For these companies, the concept of big data is not new.

However, two important trends make this era of big data quite different:

- The digitization of virtually "everything" now creates new types of large and real-time data across a broad range of industries. Much of this is non-standard data: for example, streaming, geospatial or sensor-generated data that does not fit neatly into traditional, structured, relational warehouses.
- Today's advanced analytics technologies and techniques enable organizations to extract insights from data with previously unachievable levels of sophistication, speed and accuracy.

Across industries and geographies, our study found that organizations are taking a pragmatic approach to big data. The most effective big data solutions identify business requirements first, and then tailor the infrastructure, data sources and analytics to support the business opportunity. These organizations extract new insights from existing and newly available internal sources of information, define a big data technology strategy and then incrementally upgrade their infrastructures accordingly over time.

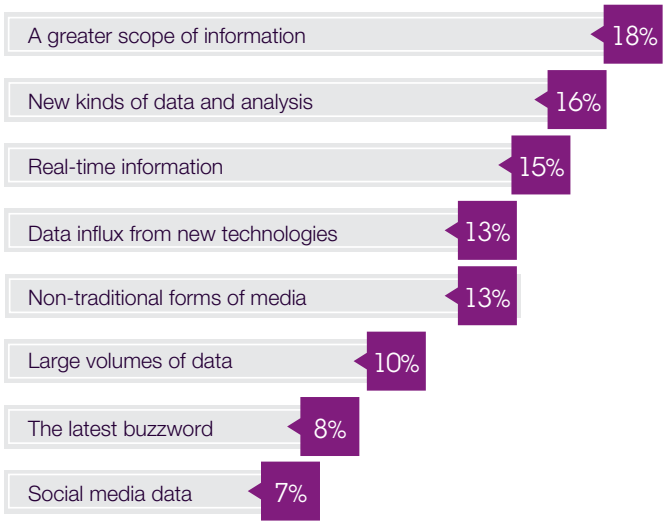
Our study findings led to five key recommendations for organizations to progress their big data efforts and seek the greatest business value from big data:

- Commit initial efforts to customer-centric outcomes
- Develop an enterprise-wide big data blueprint
- Start with existing data to achieve near-term results
- Build analytics capabilities based on business priorities
- Create a business case based on measurable outcomes.

Defining big data

Much of the confusion about big data begins with the definition itself. To understand our study respondents' definition of the term, we asked each to select up to two characteristics of big data. Rather than any single characteristic clearly dominating among the choices, respondents were divided in their views on whether big data is best described by today's greater volume of data, the new types of data and analysis, or the emerging requirements for more real-time information analysis (see Figure 1).

Defining big data



Respondents were asked to choose up to two descriptions about how their organizations view big data from the choices above. Choices have been abbreviated, and selections have been normalized to equal 100%. Total respondents=1144.

Figure 1: **Respondents were split in their views of big data.**

These results align with a useful way of characterizing three dimensions of big data – "the three Vs:" volume, variety and velocity. And while they cover the key attributes of big data itself, we believe organizations need to consider an important fourth dimension: veracity. Inclusion of veracity as the fourth big data attribute emphasizes the importance of addressing and managing for the uncertainty inherent within some types of data (see Figure 2).

The convergence of these four dimensions helps both to define and distinguish big data:

Volume: The amount of data. Perhaps the characteristic most associated with big data, volume refers to the mass quantities of data that organizations are trying to harness to improve decision-making across the enterprise. Data volumes continue to increase at an unprecedented rate. However, what constitutes truly "high" volume varies by industry and even geography, and is smaller than the petabytes and zetabytes often referenced. Just over half of respondents consider datasets between one terabyte and one petabyte to be big data, while another 30 percent simply didn't know how big "big" is for their organization. Still, all can agree that whatever is considered "high volume" today will be even higher tomorrow.

Variety: Different types of data and data sources. Variety is about managing the complexity of multiple data types, including structured, semi-structured and unstructured data. Organizations need to integrate and analyze data from a complex array of both traditional and non-traditional information sources, from within and outside the enterprise. With the explosion of sensors, smart devices and social collaboration technologies, data is being generated in countless forms, including: text, web data, tweets, sensor data, audio, video, click streams, log files and more.

Velocity: Data in motion. The speed at which data is created, processed and analyzed continues to accelerate. Contributing to higher velocity is the real-time nature of data creation, as well as the need to incorporate streaming data into business processes and decision making. Velocity impacts latency – the lag time between when data is created or captured, and when it is accessible. Today, data is continually being generated at a pace

Big data in dimensions

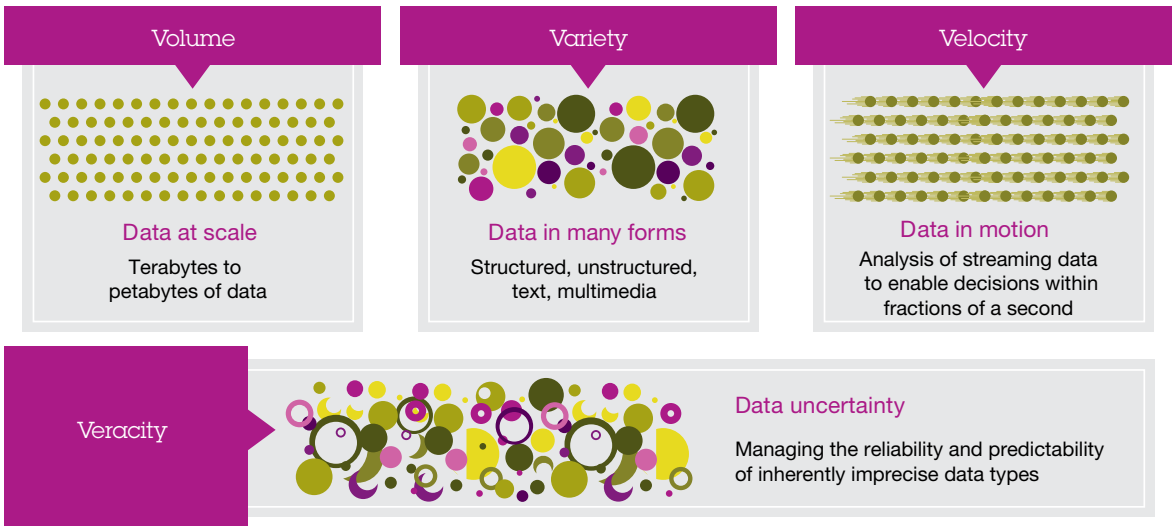


Figure 2: **Four dimensions of big data.**

that is impossible for traditional systems to capture, store and analyze. For time-sensitive processes such as real-time fraud detection or multi-channel "instant" marketing, certain types of data must be analyzed in real time to be of value to the business.

Veracity: Data uncertainty. Veracity refers to the level of reliability associated with certain types of data. Striving for high data quality is an important big data requirement and challenge, but even the best data cleansing methods cannot remove the inherent unpredictability of some data, like the weather, the economy, or a customer's actual future buying decisions. The need to acknowledge and plan for uncertainty is a dimension of big data that has been introduced as executives seek to better understand the uncertain world around them (see sidebar, "Veracity, the fourth 'V.'").[2]

Ultimately, big data is a combination of these characteristics that creates an opportunity for organizations to gain competitive advantage in today's digitized marketplace. It enables companies to transform the ways they interact with and serve their customers, and allows organizations – even entire industries – to transform themselves. Not every organization will take the same approach toward engaging and building its big data capabilities. But opportunities to utilize new big data technology and analytics to improve decision-making and performance exist in every industry.

Organizations are being practical about big data

Notwithstanding some of the hype, it is commonly agreed that we are in the early stages of enterprise big data adoption. In this study, we use the term "big data adoption" to represent a natural progression of the data, sources, technologies and skills that are necessary to create a competitive advantage in the globally integrated marketplace.

Veracity, the fourth "V"

Some data is inherently uncertain, for example: sentiment and truthfulness in humans; GPS sensors bouncing among the skyscrapers of Manhattan; weather conditions; economic factors; and the future. When dealing with these types of data, no amount of data cleansing can correct for it. Yet despite uncertainty, the data still contains valuable information. The need to acknowledge and embrace this uncertainty is a hallmark of big data.

Uncertainty manifests itself in big data in many ways. It is in the skepticism that surrounds data created in human environ-ments like social networks; in the unknowingness of how the future will unfold and of how people, nature or unseen market forces will react to the variability of the world around them.

An example of this uncertainty is in energy production: the weather is uncertain, but a utility company must still forecast production. In many countries, regulators require a percentage of production come from renewable sources, yet neither wind nor clouds can be forecast with precision. So how do you plan?

To manage uncertainty, analysts need to create context around the data. One way to achieve this is through data fusion, where combining multiple less reliable sources creates a more accurate and useful data point, such as social comments appended to geospatial location information. Another way to manage uncertainty is through advanced mathematics that embraces it, such as robust optimization techniques and fuzzy logic approaches.

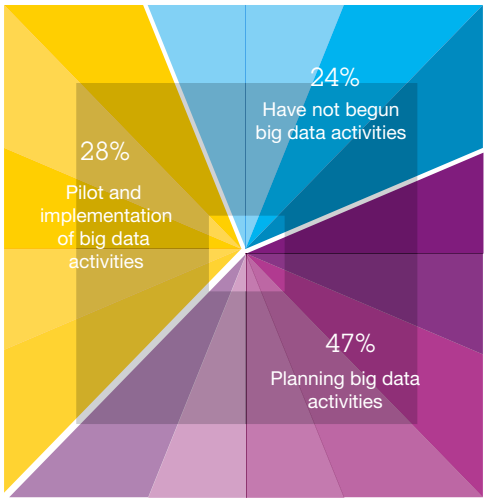
Humans, by nature, dislike uncertainty, but just ignoring it can create even more problems than the uncertainty itself. In the era of big data, executives will need to approach the dimension of uncertainty differently. They will need to acknowledge it, embrace it and determine how to use it to their advantage; the one certainty about uncertainty is that it is not likely to go away.

Our Big Data @ Work survey confirms that most organizations are currently in the early stages of big data development efforts, with the majority focused either on understanding the concepts (24 percent) or defining a roadmap related to big data (47 percent). However, 28 percent of respondents are in leading-edge organizations where they are developing proofs of concepts (POCs) or have already implemented big data solutions at scale (see Figure 3).

By analyzing survey response, five key study findings show some common and interesting trends and insights:

- Across industries, the business case for big data is strongly focused on addressing customer-centric objectives
- A scalable and extensible information management foundation is a prerequisite for big data advancement
- Organizations are beginning their pilots and implementations by using existing and newly accessible internal sources of data
- Advanced analytic capabilities are required, yet often lacking, for organizations to get the most value from big data
- As organizations' awareness and involvement in big data grows, we see four stages of big data adoption emerging.

Big data activity



Respondents were asked to identify the current state of big data activities within their organizations. Percentage does not equal 100% due to rounding. Total respondents=1061

Figure 3: **Most organizations are in early stages of big data development efforts.**

Customer analytics are driving big data initiatives

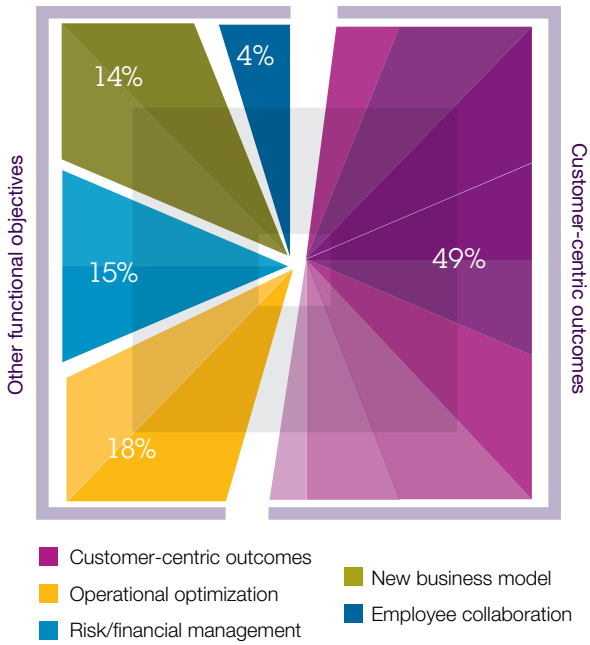
When asked to rank their top three objectives for big data, nearly half of the respondents identified customer-centric objectives as their organization's top priority (see Figure 4). Organizations are committed to improving the customer experience and better understanding customer preferences and behavior. Understanding today's "empowered consumer" was also identified as a high priority in both the 2011 IBM Global Chief Marketing Officer Study and 2012 IBM Global Chief Executive Officer Study.[3]

Companies clearly see big data as providing the ability to better understand and predict customer behaviors, and by doing so, improve the customer experience. Transactions, multi-channel interactions, social media, syndicated data through sources like loyalty cards, and other customer-related information have increased the ability of organizations to create a complete picture of customers' preferences and demands – a goal of marketing, sales and customer service for decades.

Through this deeper understanding, organizations of all types are finding new ways to engage with existing and potential customers. This principle clearly applies in retail, but equally as well in telecommunications, healthcare, government, banking and finance, and consumer products where end-consumers and citizens are involved, and in business-to-business interactions among partners and suppliers.

In fact, big data can be a two-way street between customers and organizations. For example, the Ford Focus Electric car produces vast amounts of data while being driven and when it is parked. While in motion, the driver is constantly updated with information about the vehicle's acceleration, braking, battery charge and location.[4] This is useful for the driver, but the data is also streamed back to Ford engineers who learn about customers' driving habits including how, when and where they charge their cars.[5] And while the vehicle is at rest, it continues to stream data about the car's tire pressure and battery system to the nearest smart phone.[6]

Business-driven outcomes



Respondents were asked to rank their top functional objectives for big data within their organizations. Responses were weighted and aggregated. Total respondents=1067

Figure 4: **Nearly half of respondents' big data efforts target customer-centric outcomes.**

Big data allows a more complete picture of customers' preferences and demands; through this deeper understanding, organizations of all types are finding new ways to engage with existing and potential customers.

Multiple benefits stem from this customer-focused scenario as big data enables valuable new kinds of collaboration. Drivers get useful, up-to-the-second information while engineers back in Detroit aggregate the information about driving behaviors to gain customer insights and plan product improvements. What's more, utilities and other third-party vendors analyze millions of miles' worth of driving data to decide where to locate new charging stations, and how to protect the fragile utility grids from overloading.[7]

Organizations worldwide are serving their customers better and improving operations through big data. Companies like Mcleod Russel India Limited completely eliminated systems downtime in the tea trade through more accurate tracking of the harvest, production and marketing of up to 100 million kilos of tea each year.[8] Premier Healthcare Alliance used enhanced data sharing and analytics to improve patient outcomes while reducing spending by US$2.85 billion.[9] And Santam improved the customer experience by implementing predictive analytics to reduce fraud (see sidebar, "Santam: Predictive analytics improve fraud detection and speed up claims processing").

In addition to customer-centric objectives, other functional objectives are also being addressed through early applications of big data. Operational optimization, for example, was cited by 18 percent of respondents, but consists largely of pilot projects. Other big data applications that they frequently mentioned include: risk/financial management, employee collaboration and enabling new business models.

Santam: Predictive analytics improve fraud detection and speed up claims processing[10]

Fraud is a very real challenge for insurance companies around the world. Whether fraud is on a large scale, such as arson, or involves a smaller claim such as an inflated auto repair bill, payouts for fraudulent claims cost companies millions of dollars every year – and that cost gets passed down to the customer in the form of higher insurance premiums. Insurance companies are fighting fraud, but traditional techniques such as legal action and private investigation are time consuming and cost prohibitive.

As South Africa's largest short-term insurance provider, Santam definitely felt the sting of insurance fraud. Fraud losses accounted for 6 to 10 percent of annual premium costs for Santam customers. And fraud had another consequence – poor operational efficiency. Because agents had to handle and investigate both high- and low-risk claims, all claims took a minimum of three days to settle, and Santam began to feel its good reputation for customer service suffer in the age where customers demand fast results.

Santam gained the ability to catch fraud early with an advanced analytics solution that captures data from incoming claims, assesses each claim against identified risk factors and segments claims to five risk categories – separating likely fraudulent claims and higher-risk from low-risk cases. With the new system, the company not only saves millions previously lost to insurance fraud, but also drastically reduces processing time for low-risk claims, leading to resolution in less than an hour for some customers. In the first few months after implementation, Santam also uncovered a major auto insurance fraud syndicate. Big data, predictive analytics and risk segmentation helped the company identify patterns that led to the fraud detection.

Big data is dependent upon a scalable and extensible information foundation

The promise of achieving significant, measurable business value from big data can only be realized if organizations put into place an information foundation that supports the rapidly growing volume, variety and velocity of data. We asked respondents to identify the current state of their big data infrastructures. Almost two-thirds report having started their big data journeys with an information foundation that is integrated, scalable, extensible and secure. Four information management components were cited most often as part of respondents' big data initiatives (see Figure 5).

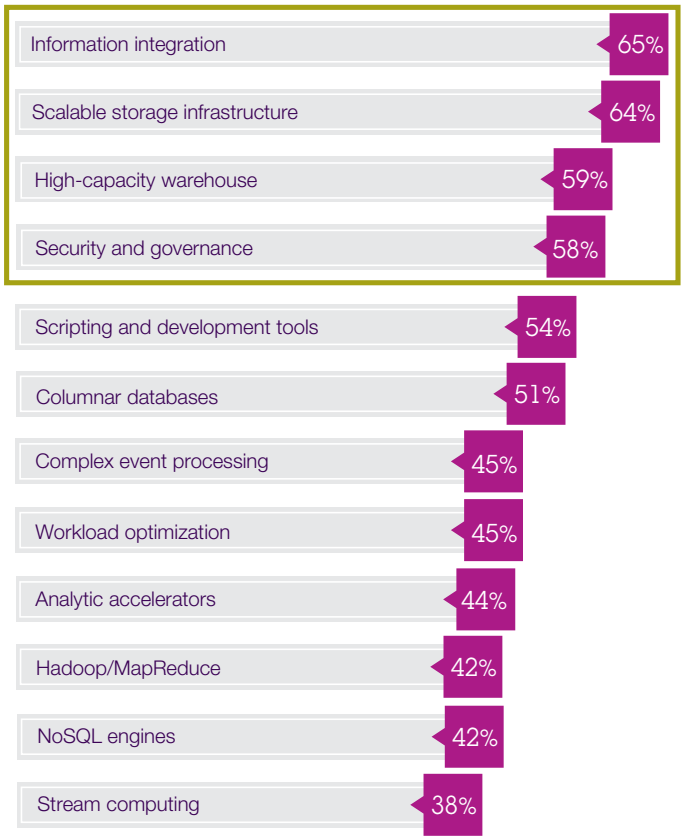
Integrated information is a core component of any analytics effort, and it is even more important with big data. As noted in the 2011 IBM Institute for Business Value study on advanced analytics, an organization's data has to be readily available and accessible to the people and systems that need it.[11]

Master data management and the integration of key data types – customer, product, vendor, employee and the like – require cross-enterprise data that is governed according to a single enterprise standard. The inability to connect data across organizational and department silos has been a business intelligence challenge for years. This integration is even more important, yet much more complex, with big data. Among respondent organizations with active big data efforts, 65 percent consider their integrated information capability to be sufficient to support big data.

The next most prevalent information management foundation components in big data initiatives are a scalable storage infrastructure and high-capacity warehouse. Each supports the rapid growth of current and future data coming into the organization.

On the surface, a combination of adding storage and one or more larger servers can support the growth of an information management foundation. However, it is important to understand that anticipating and architecting the infrastructure is key to delivering the business value of the intended business case. Organizations need to consider how best to support the ebb and flow of data to enable users to access data when needed, as well as how data can be analyzed within the business's time constraints (whether days, hours, seconds, milliseconds). This balanced configuration and deployment of servers and storage results in a more optimized infrastructure.

Big data infrastructure



Respondents with active big data efforts were asked which platform components are currently either in pilot or integrated into the architecture. Each data point was collected independently. Total respondents for each data point range from 297 to 351.

Figure 5: **Components of respondents' big data infrastructures.**

These technologies also manage the increasing velocity of inbound – and stored – data by enabling consistent, automated data movement across the enterprise as more people require access to additional – and different – types of information. Emerging technologies such as data tiering and compression, and scale-out file systems, along with in-memory databases, are enabling the management of much larger workloads than conventional warehouses. For many organizations, improving the capability to manage growing volumes is the first big data priority, followed closely by addressing the expanding variety of data. (See sidebar, "Vestas: Better data analysis capabilities lower costs and improve effectiveness")

Strong security and governance processes are in place at 58 percent of the organizations who report having active big data efforts underway. While security and governance have long been an inherent part of business intelligence, the added legal, ethical and regulatory considerations of big data introduce new risks and expand the potential for very public missteps, as we have already seen in some companies that have lost control of data or use it in questionable ways.

As a result, data security – and especially data privacy – is a critical part of information management, according to several interviewed subject matter experts and business executives. Security and governance will become even more important and daunting as organizations embrace new sources of information, especially social media data. Compounding this challenge, privacy regulations are still evolving and can vary greatly by country.

"There is the perception that privacy and security is easy, but it's very regulated, very closely watched," one telecommunications industry executive explained. And it's not just governmental agencies watching, but also the customers themselves. The executive continued: "There are a number of new areas – like web browsing data – where a gray area exists between what's legal and what's right. We have taken the approach to consider every action using the standard of what the customer would think if (the way we used the data) was splashed across the front page."

Vestas: Better data analysis capabilities lower costs and improve effectiveness[12]

Wind turbines are a multimillion dollar investment with a typical lifespan of 20 to 30 years. To determine the optimal placement for a turbine, a large number of location-dependent factors must be considered, including temperature, precipitation, wind velocity, humidity and atmospheric pressure.

For Vestas Wind Systems A/S (Vestas), a Danish wind turbine producer, the data analysis process used to create its customer turbine location models was becoming increasingly unsatisfactory; the process took several weeks to execute and could not support analysis of the desired very large set of data the company deemed necessary for precision turbine placement and power forecasting. Vestas engineers wanted to start developing their own forecasts using actual recorded data for existing customer turbines rather than industry models; the challenge was to increase data capacity requirements to a projected six petabytes.

Using a big data solution on a supercomputer that is one of the world's largest to-date and a modeling solution designed to harvest insights from an expanded set of factors including both structured and unstructured data, the company can now help its customers optimize turbine placement and, as a result, turbine performance.

This new information environment enables the company to manage and analyze weather and location data in ways that were previously not possible to gain insights that can lead to improved decisions for wind turbine placement and operations, as well as more accurate power production forecasts. The detailed models mean greater business case certainty, quicker results, and increased predictability and reliability, which decreases cost to customers per kilowatt hour produced and increases the precision of customer ROI estimates. These technologies reduce by approximately 97 percent – from weeks to hours – the response time for business user requests, and greatly improve the effectiveness of turbine placement.

The costs associated with upgrading infrastructures were raised as a concern by several interviewed executives. Senior leadership, they reported, require a solid, quantifiable business case, one that defines incremental investments along with opportunities to rationalize and optimize the costs of their information management environments. Lower-cost architectures – including cloud computing, strategic outsourcing and value-based pricing – were cited as tactics being deployed. Yet others invested in their information platforms based on the conviction that the business opportunity was worth the associated incremental costs.

Initial big data efforts are focused on gaining insights from existing and new sources of internal data

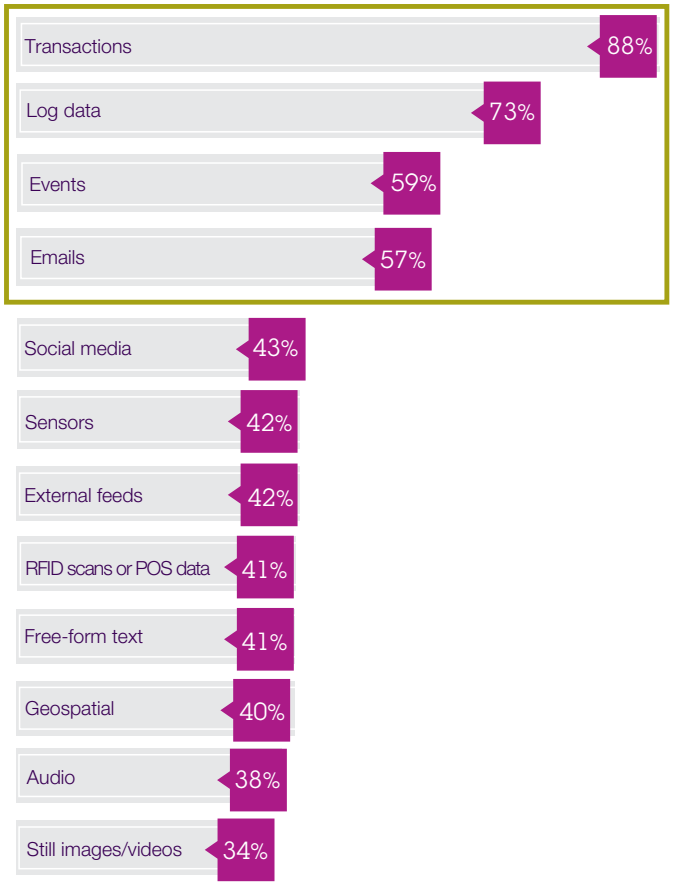
Most early big data efforts are targeted at sourcing and analyzing internal data. According to our survey, more than half of the respondents reported internal data as the primary source of big data within their organizations. This suggests that companies are taking a pragmatic approach to adopting big data and also that there is tremendous untapped value still locked away in these internal systems (see Figure 6).

As expected, internal data is the most mature, well-understood data available to organizations. It has been collected, integrated, structured and standardized through years of enterprise resource planning, master data management, business intelligence and other related work. By applying analytics, internal data extracted from customer transactions, interactions, events and emails can provide valuable insights (see sidebar, "Automercados Plaza's: Greater revenue through greater insight"). However, in many organizations, the size and scope of this internal data, such as detailed transactions and operational log data, have become too large or varied to manage within traditional systems.

Almost three out of four respondents with active big data efforts are analyzing log data. This is "machine/sensor generated" data produced to record the details of automated functions performed within business or information systems – data that has outgrown the ability to be stored and analyzed by many traditional systems. As a result, much of this data is collected, but not analyzed.

Executive interviews confirmed that many CIOs who are guiding their companies' big data initiatives are beginning with these untapped sources of internal information, using the additional processing power provided by a more scalable infrastructure.

Big data sources



Respondents with active big data efforts were asked which data sources they currently collect and analyze. Each data point was collected independently. Total respondents for each data point range from 557 to 867.

Figure 6: **Organizations are mainly using internal data sources for big data efforts.**

Automercados Plaza's: Greater revenue through greater insight[13]

Automercados Plaza's, a family-owned chain of grocery stores in Venezuela, found itself with more than six terabytes of product and customer data spread across different systems and databases. As a result, it could not easily assess operations at each store and executives knew there were valuable insights to be found.

"We had a big mess related to pricing, inventory, sales, distribution and merchandising," says Jesus Romero, CIO, Automercados Plaza's. "We have nearly US$20 million in inventory and we tracked related information in different systems and compiled it manually. We needed an integrated view to understand exactly what we have."

By integrating information across the enterprise, the grocery chain has realized a nearly 30 percent increase in revenue and a US$7 million increase in annual profitability. Mr. Romero attributes these increases to better inventory management and the ability to more quickly adjust to changing market conditions. For example, the company has prevented losses for about 35 percent of its products now that it can schedule price reductions to sell perishable products before they spoil.

Big data requires strong analytics capabilities

Big data does not create value, however, until it is put to use to solve important business challenges. This requires access to more and different kinds of data, as well as strong analytics capabilities that include both software tools and the requisite skills to use them.

Examining those organizations engaged in big data activities reveals that they start with a strong core of analytics capabilities designed to address structured data. Next, they add capabilities to take advantage of the wealth of data coming into the organization that is both semi-structured (data that can be converted to standard data forms) and unstructured (data in non-standard forms).

More than 75 percent of respondents with active big data efforts reported using core analytics capabilities, such as query and reporting, and data mining to analyze big data, while more than 67 percent report using predictive modeling. Beginning with these foundational analytics capabilities is a pragmatic way to start interpreting and analyzing big data, especially when it is being stored in a relational database. (See Figure 7).

The need for more advanced data visualization capabilities increases with the introduction of big data. Datasets are often too large for business or data analysts to view and analyze with traditional reporting and data mining tools. In our study, respondents said that 71 percent of active big data efforts rely on data visualization skills.

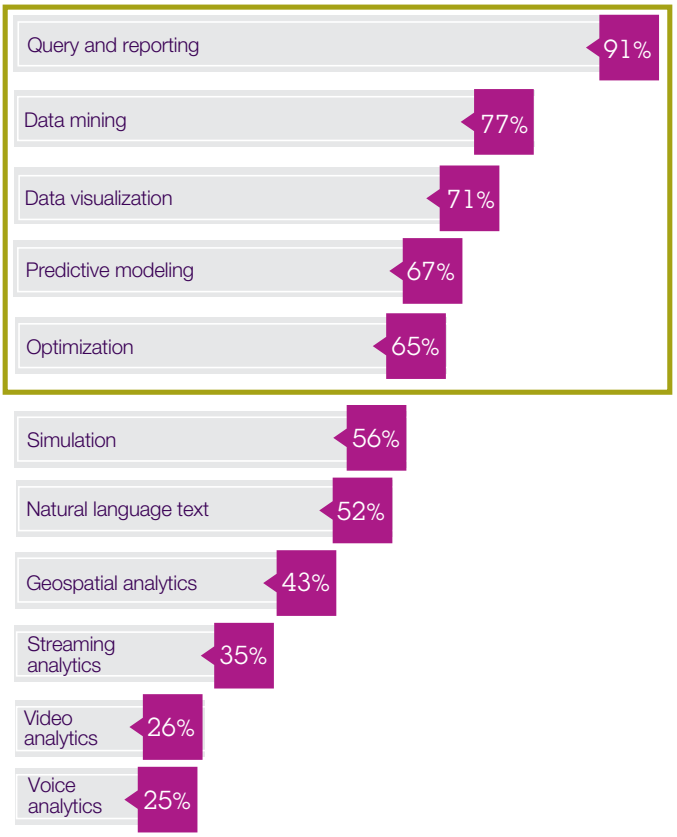
Organizations engaged in big data require increasingly more advanced capabilities to find patterns in the inherent complexity. To accomplish this, respondents are applying optimization models and advanced analytics to better understand how to transform key business processes. They are using simulation capabilities to analyze the myriad of variables available within big data. Our survey found that more than 50 percent of active big data efforts are using these advanced modeling capabilities.

Today, most companies are directing their initial big data focus toward analyzing structured data. But big data also creates the need to analyze multiple data types, including a variety of types that may be entirely new for many organizations. In more than half of the active big data efforts, respondents reported using advanced capabilities designed to analyze text in its natural state, such as the transcripts of call center conversations. These analytics include the ability to interpret and understand the nuances of language, such as sentiment, slang and intentions.

Having the capabilities to analyze unstructured (for example, geospatial location data, voice and video) or streaming data continues to be a challenge for most organizations. While the hardware and software in these areas are maturing, the skills are in short supply. Fewer than 25 percent of respondents with active big data efforts reported having the required capabilities to analyze extremely unstructured data like voice and video.

Acquiring or developing these more advanced technical and analytic capabilities required for big data advancement is becoming a top challenge among many organizations with active big data efforts. Among these organizations, the lack of advanced analytical skills is a major inhibitor to getting the most value from big data.

Big data analytics capabilities



Respondents with active big data efforts were asked which analytics capabilities are currently available within their organizations. Each data point was collected independently. Total respondents for each data point range from 508 to 870.

Figure 7: **Respondents are applying a variety of advanced analytics.**

The emerging pattern of big data adoption is focused upon delivering measurable business value

To better understand the big data landscape, we asked respondents to describe the level of big data activities in their organizations today. The results suggest four main stages of big data adoption and progression along a continuum that we have labeled Educate, Explore, Engage and Execute (see Figure 8).

Educate: Building a base of knowledge *(24 percent of respondents)*

In the Educate stage, the primary focus is on awareness and knowledge development. Almost 25 percent of respondents indicated they are not yet using big data within their organizations. While some remain relatively unaware of the topic of big data, our interviews suggest that most organizations in this stage are studying the potential benefits of big data technologies and analytics, and trying to better understand how big data can help address important business opportunities in their own industries or markets. Within these organizations, it is mainly individuals doing the knowledge gathering as opposed to formal work groups, and their learnings are not yet being used by the organization. As a result, the potential for big data has not yet been fully understood and embraced by the business executives.

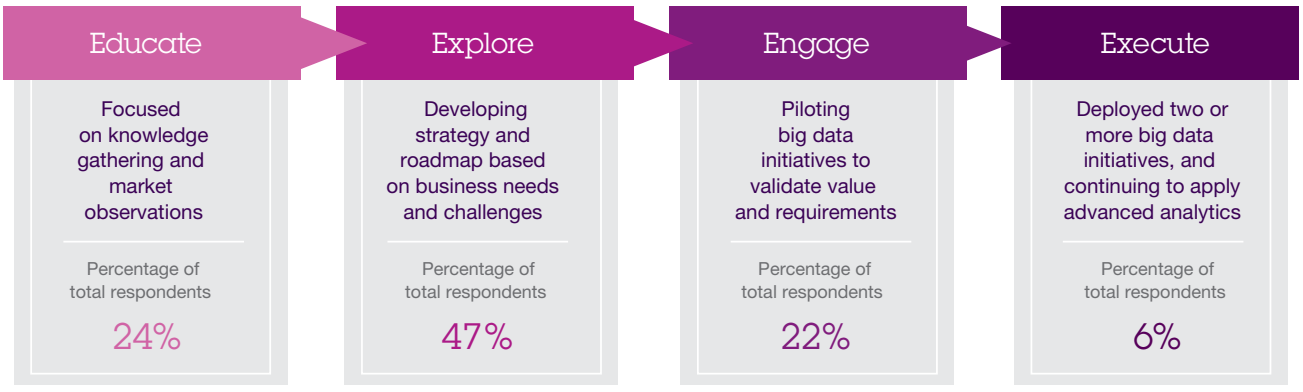
Explore: Defining the business case and roadmap *(47 percent)*

The focus of the Explore stage is to develop an organization's roadmap for big data development. Almost half of respondents reported formal, ongoing discussions within their organizations about how to use big data to solve important business challenges. Key objectives of these organizations include developing a quantifiable business case and creating a big data blueprint. This strategy and roadmap takes into consideration existing data, technology and skills, and then outlines where to start and how to develop a plan aligned with the organization's business strategy.

Engage: Embracing big data *(22 percent)*

In the Engage stage, organizations begin to prove the business value of big data, as well as perform an assessment of their technologies and skills. More than one in five respondent organizations is currently developing proofs-of-concept (POCs) to validate the requirements associated with implementing big data initiatives, as well as to articulate the expected returns. Organizations in this group are working – within a defined, limited scope – to understand and test the technologies and skills required to capitalize on new sources of data.

Big data adoption stages



Respondents were asked to identify the current state of big data activities within their organizations. Percentage does not equal 100% due to rounding. Total respondents=1061

Figure 8: **Four stages in the emerging pattern of big data adoption.**

Execute: Implementing big data at scale *(6 percent)*
In the Execute stage, big data and analytics capabilities are more widely operationalized and implemented within the organization. However, only 6 percent of respondents reported that their organizations have implemented two or more big data solutions at scale – the threshold for advancing to this stage. The small number of organizations in the Execute stage is consistent with the implementations we see in the marketplace. Importantly, these leading organizations are leveraging big data to transform their businesses and thus are deriving the greatest value from their information assets. With the rate of enterprise big data adoption accelerating rapidly – as evidenced by 22 percent of respondents in the Engage stage, with either POCs or active pilots underway – we expect the percentage of organizations at this stage to more than double over the next year.

More changes by stage as big data capabilities evolve

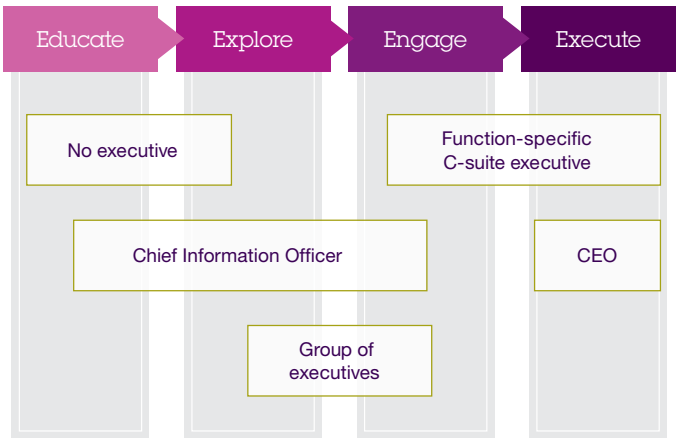
In addition to the five key study findings, our analysis revealed several other stage-related patterns associated to executive sponsorship, data availability requirements and primary obstacles. These patterns show the interconnected series of steps that organizations are taking to evolve their big data infrastructures and analytics capabilities.

Big data sponsorship

A deeper examination of big data adoption illustrates an interesting pattern of changing sponsorship (see Figure 9). Across our sample, respondents report that more than one-third of big data efforts are being driven by the CIO. However, the majority of CIO-driven efforts are in the early stages of adoption as the organization is investing in technology and beginning to identify business opportunities and requirements.

As organizations advance to later stages, sponsorship is being provided by a business executive – either a function-specific executive such as a CMO or CFO, or even the CEO. Notably, this sponsorship model with a single point-of-focus by a business executive is viewed as critical to big data success.

Big data sponsorship



Respondents were asked who is most closely associated with the mandate for the use of big data and analytics insights. Box placement reflects the degree to which an executive is dominant in each stage.
Total respondents=1028

Figure 9: **Leadership shifts are evident as big data efforts progress within an organization.**

This pattern suggests that organizations may initially focus on technology and building out their big data infrastructures, but as they begin to develop their business cases and roadmaps, the sponsorship shifts to one or more business executives. Even so, the CIO and the IT organization must continue to play an important role in executing the agreed-upon roadmap.

Data availability

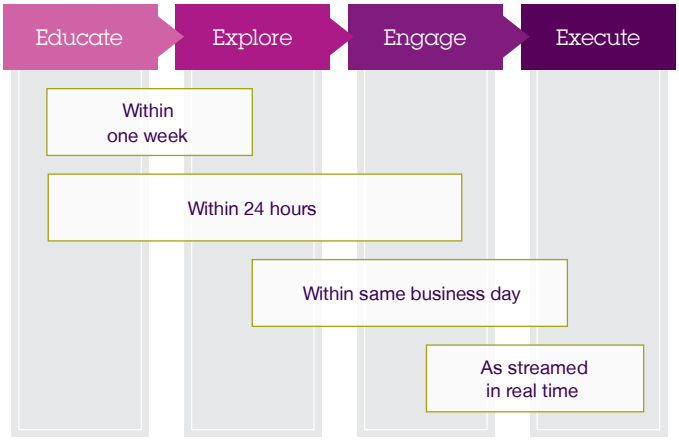
As depicted in Figure 10, we see how data availability requirements change dramatically as companies mature their big data efforts. Analysis of responses revealed that no matter the stage of big data adoption, organizations face increasing demands to reduce the latency from data capture to action. Executives, it seems, are increasingly considering the value of timely data in making strategic and day-to-day business decisions. Data is no longer just something that supports a decision; it is a mission-critical component in making that decision.

We anticipate that the demands for more real-time access will continue to rise as business models evolve and organizations invest in technologies required for streaming data, in-memory analytics, machine-to-machine processing and other innovative advancements.

Big data obstacles

Challenges that inhibit big data adoption differ as organizations move through each of the big data adoption stages. But our findings show one consistent challenge – regardless of stage – and that is the ability to articulate a compelling business case (see Figure 11). At every stage, big data efforts come under fiscal scrutiny. The current global economic landscape has left businesses with little appetite for new technology investments without measurable benefits – a requirement that, of course, is not exclusive to big data initiatives. After organizations successfully implement POCs, the biggest challenge becomes finding the skills to operationalize big data, including: technical, analytical and governance skills.

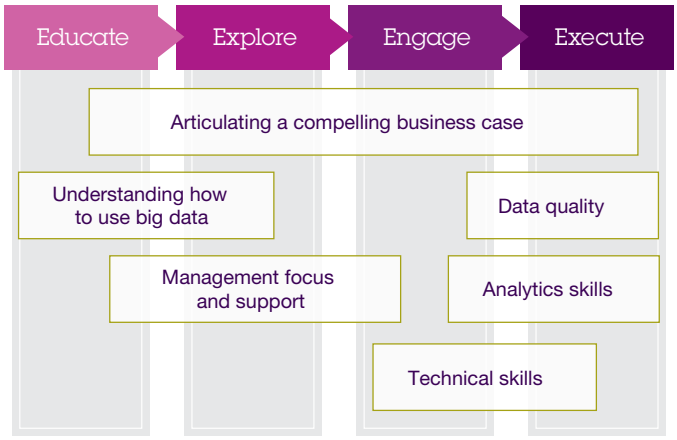
Required data availability



Respondents were asked how quickly data is required to be available for business users or processes. Box placement reflects the prevalence of that requirement within a stage. Total respondents=973

Figure 10: **Stepping up to the challenge of real-time data.**

Primary obstacles



Respondents were asked to rank the top challenges for big data within their organizations. Box placement reflects the prevalence of that challenge in each stage. Responses were weighted and aggregated. Total respondents=1062

Figure 11: **Understanding key big data adoption obstacles.**

Recommendations: Cultivating big data adoption

IBM analysis of our Big Data @ Work Study findings provided new insights into how organizations at each stage are advancing their big data efforts. Driven by the need to solve business challenges, in light of both advancing technologies and the changing nature of data, organizations are starting to look closer at big data's potential benefits. To extract more value from big data, we offer a broad set of recommendations to organizations as they proceed down the path of big data.

Commit initial efforts to customer-centric outcomes

It is imperative that organizations focus big data initiatives on areas that can provide the most value to the business. For many industries, this will mean beginning with customer analytics that enable better service to customers as a result of being able to truly understand customer needs and anticipate future behaviors.

Mass digitization, one of the forces that helped to create the surge in big data, has also changed the balance of power between the individual and the institution. If organizations are to understand and provide value to empowered customers and citizens, they have to concentrate on getting to know their customers as individuals. They will also need to invest in new technologies and advanced analytics to gain better insights into individual customer interactions and preferences.

But today's customers – end consumers or business-to-business customers – want more than just understanding. To effectively cultivate meaningful relationships with their customers, organizations must connect with them in ways their customers perceive as valuable.

The value may come through more timely, informed or relevant interactions; it may also come as organizations improve the underlying operations in ways that enhance the overall experience of those interactions. Either way, analytics fuels the insights from big data that are increasingly becoming essential to creating that level of depth in these relationships.

Develop an enterprise-wide big data blueprint

A blueprint encompasses the vision, strategy and requirements for big data within an organization, and is critical to establishing alignment between the needs of business users and the implementation roadmap of IT. It creates a common understanding of how the enterprise intends to use big data to improve its business objectives.

An effective blueprint defines the scope of big data within the organization by identifying the key business challenges to which it will be applied, the business process requirements that define how big data will be used, and the architecture which includes the data, tools and hardware needed to achieve it. It is the basis for developing a roadmap to guide the organization through a pragmatic approach to develop and implement its big data solutions in ways that create sustainable business value.

Start with existing data to achieve near-term results

To achieve near-term results while building the momentum and expertise to sustain a big data program, it is critical that companies take a pragmatic approach. As respondents confirmed, the most logical and cost-effective place to start looking for new insights is within the enterprise.

Looking internally first allows organizations to leverage their existing data, software and skills, and to deliver near-term business value and gain important experience as they then consider extending existing capabilities to address more complex sources and types of data. Most organizations will want to do this to take advantage of the information stored in existing repositories while scaling their data warehouse(s) to handle larger volumes and varieties of data.

Build analytics capabilities based on business priorities

Throughout the world, organizations face a growing variety of analytics tools while also facing a critical shortage of analytical skills. Big data effectiveness hinges on addressing this significant gap. In short, organizations will have to invest in acquiring both tools and skills. As part of this process, it is expected that new roles and career models will emerge for individuals with the requisite balance of analytical, functional and IT skills.

Attention to the professional development and career progression of in-house analysts – who are already familiar with the organization's unique business processes and challenges – should be a top priority for business executives. At the same time universities and individuals themselves, regardless of background or specialty, have an obligation to build solid analytical skills.

Create a business case based on measurable outcomes

To develop a comprehensive and viable big data strategy and the subsequent roadmap requires a solid, quantifiable business case. Therefore, it is important to have the active involvement and

sponsorship from one or more business executives throughout this process. Equally important to achieving long-term success is strong, ongoing business and IT collaboration.

Many organizations are basing their business cases on the following benefits that can be derived from big data:

- *Smarter decisions* – Leverage new sources of data to improve the quality of decision making.
- *Faster decisions* – Enable more real-time data capture and analysis to support decision making at the "point of impact," such as when a customer is navigating your website or on the telephone with a customer service representative.
- *Decisions that make a difference* – Focus big data efforts toward areas that provide true differentiation.

An important principle underlies each of these recommendations: business and IT professionals must work together throughout the big data journey. The most effective big data solutions identify the business requirements first, and then tailor the infrastructure, data sources and quantitative analysis to support that business opportunity.

Additional recommendations by stage: Start where you are

Certain key activities are characteristic of each stage in the big data adoption lifecycle. The following recommendations by stage offer a proven and practical approach for moving from one stage to the next.

Educate to Explore: Create a foundation for action

- Continue to expand your knowledge by focusing on use cases where big data is providing competitive advantage to organizations, both inside and outside of your Industry.
- Work with different business units and functions to identify your most critical business opportunities and challenges that can be addressed with better and more timely information access. Many organizations begin with customer data and analytics to support their front-office transformation agenda.
- Focus on strengthening your information management environment and infrastructure, including the development of a big data blueprint. These blueprints are often based on industry standards, reference architectures and other available technical frameworks and resources.

Explore to Engage: Put plans into action

- Confirm active business leader sponsorship as you develop your big data strategy and roadmap.
- Develop the business case for one or two key business opportunities or challenges that you plan to address through POCs or pilot project(s).
- While beginning to plan for longer-term requirements, regularly confirm that your information management foundation and IT infrastructure are able to support the big data technologies and capabilities required for the POC or pilot.
- Assess your current information governance processes and their readiness to address the new aspects of big data.
- Analyze existing skill sets of internal resources, and begin gap analysis of where you need to grow and/or hire additional skills.

Engage to Execute: Understand the opportunities and challenges ahead

- Actively promote pilot project successes to sustain momentum while beginning to engage other parts of the business.
- Finalize the business case with the validation and quantification of projected returns on investments and benefits, including defined success criteria and metrics.
- Identify the business process modifications and improvements expected from having access to better and more timely information (for example, marketing, sales, customer service and social media sites).
- Develop a competency plan to confirm the availability of adequate technical and quantitative skills that are required to achieve short-term and longer-term objectives.
- Document the detailed project plan for migrating pilot(s) into production. This plan should include confirmation of expected business value, costs, resources and project timelines.

Execute stage: Embrace the innovation of big data

- Document quantifiable outcomes of early successes to bolster future efforts.
- Initiate formal big data communications across the organization to continue building support and momentum.
- Focus on extending technologies and skills required to address new big data challenges across business units, functions and geographies.
- Remain vigilant about information governance (including information lifecycle management), privacy and security.
- Continue to evaluate rapidly-evolving big data tools and technology. Balance existing infrastructure with newer technologies that increase scalability, optimization and resiliency.

Kick off your big data evolution

To compete in a globally-integrated economy, it is increasingly clear that today's organizations need a comprehensive understanding of markets, customers, products, regulations, competitors, suppliers, employees and more. This understanding demands the effective use of information and analytics. In fact, next to their employees, many companies consider information to be their most valuable and differentiated asset.

Now, with the emergence and expanding adoption of big data, organizations worldwide are discovering entirely new ways to compete and win. They are transforming themselves to take advantage of the vast array of information that is available to improve decision-making and performance throughout the enterprise. A comparatively small group of pioneering organizations is already accomplishing this by arming employees – from the C-suite to marketing to the shop floor – with the information, skills and tools required to make better and more timely decisions at the "point of impact."

Not every organization will need to manage for the full spectrum of big data capabilities. But the opportunity to utilize new data, technology and analytics exists to some degree within every industry. Organizations realize value by analyzing the volume, velocity and variety of new and existing data, and putting the right skills and tools in place to better understand their operations, customers and the marketplace as a whole. Whatever the starting point, organizations around the world will continue to expand the use of big data to gain business value and competitive advantage in today's globally integrated economy.

To learn more about this IBM Institute for Business Value study, please contact us at iibv@us.ibm.com. For a full catalog of our research, visit: **ibm.com/**iibv

Subscribe to IdeaWatch, our monthly e-newsletter featuring the latest executive reports based on IBM Institute for Business Value research: **ibm.com/**gbs/ideawatch/subscribe

Access IBM Institute for Business Value executive reports on your tablet by downloading the free "IBM IBV" app for iPad or Android.

To learn more about the Säid Business School at the University of Oxford, please visit *www.sbs.ox.ac.uk*

Related publications

Kiron, David, Rebecca Shockley, Nina Kruschwitz, Glenn Finch and Dr. Michael Haydock, *"Analytics: The widening divide: How companies are achieving competitive advantage through analytics"* IBM Institute for Business Value in collaboration with MIT Sloan Management Review. October 2011. *http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-analytics-widening-divide.html* © 2011 Massachusetts Institute for Technology.

LaValle, Steve, Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz. *"Analytics: The new path to value: How the smartest organizations are embedding analytics to transform insights into action."* IBM Institute for Business Value in collaboration with MIT Sloan Management Review. October 2010. *http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-embedding-analytics.html* © 2010 Massachusetts Institute for Technology.

Teerlink, Dr. Marc and Dr. Michael Haycock. *"Customer analytics pay off: Driving top-line growth by bringing science to the art of marketing."* IBM Institute for Business Value. September 2011. *http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-customer-analytics.html*

Authors

Michael Schroeck is a Partner and Vice President for IBM Global Business Services, where he serves as the firm's Global Information Management Foundation Leader. Michael is also an IBM Distinguished Engineer, and can be reached at *mike.schroeck@us.ibm.com*

Rebecca Shockley is the Business Analytics and Optimization Global Research Leader for the IBM Institute for Business Value, where she conducts fact-based research on the topic of business analytics to develop thought leadership for senior executives. Rebecca can be contacted at *rshock@us.ibm.com*

Dr. Janet Smart is a member and co-director of the Complex Agent-Based Dynamic Networks research cluster in the Saïd Business School, combining expertise in complex systems, systems engineering, big data and project management. Dr. Smart also works with the ATLAS project at CERN on project management and systems engineering in Big Science projects. She teaches Systems Engineering on the MSc in Major Programme Management and on the UK's Major Projects Leadership Academy programme. Janet can be reached at *Janet.Smart@sbs.ox.ac.uk*

Professor Dolores Romero-Morales is a Professor of Operations Research at Saïd Business School. She teaches the core course on Decision and Data Analytics in the MBA and EMBA programmes and has published dozens of research articles in leading journals in the areas of Supply Chain Optimisation, Data Mining and Revenue Management. She has also worked on these topics with practitioners in a range of industries. Dolores can be reached at *Dolores.Romero-Morales@sbs.ox.ac.uk*

Professor Peter Tufano is Peter Moores Dean at Saïd Business School. Professor Tufano's research focuses on innovation, especially how innovations can improve the delivery of financial services to families. Before Oxford, Professor Tufano spent 33 years at Harvard and founded D2DFund *(www.d2dfund.org).* Peter can be contacted through his office: *claire.eggleton@sbs.ox.ac.uk*

Contributors

Fred Balboni, Global Leader, Business Analytics and Optimization, IBM Global Business Services

Dr. Stephen Buckley, BAO Applied Research Leader, IBM Research

Wendy Olivier, Information Management Foundation Program Director, IBM Global Business Services

Katharyn White, Vice President, Marketing, IBM Global Business Services

The authors also acknowledge the following IBM and University of Oxford colleagues for their help in developing this executive report: Dr. John Bell. Min Chen, Michael Coleman, Richard Cuthbertson, Tom Deutsch, Angela Finley, Mark Graham, Larry Gosselin, Tina Groves, Bambi Grundwerg, Pamela Hartigan, Bernie Hogan, Matin Jouzdani, Jim Kocis, Eric Lesser, Monica Logan, Eduardo Lopez, Dr. Robin Lougee, Piyush Malik, Helen Margetts, Joni McDonald, Brian Morris, Richard Perret, Nancy Puccinelli, Tarun Ramadorai, Dean Ranalli, Eric Sall, Ralph Schroeder, Stephanie Schneider, Ronald Shelby, Neil Shephard, James Taylor, Owen Tebbutt and Andy Twigg.

References

1 LaValle, Steve, Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz. *"Analytics: The new path to value: How the smartest organizations are embedding analytics to transform insights into action."* IBM Institute for Business Value in collaboration with MIT Sloan Management Review. October 2010. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-embedding-analytics.html © 2010 Massachusetts Institute for Technology.

2 "*2012 IBM Global Technology Outlook.*" IBM Research. March 2012. http://www.research.ibm.com/files/pdfs/gto_booklet_executive_review_march_12.pdf

3 *"From Stretched to Strengthened: Insights from the IBM Chief Marketing Officer Study."* IBM Institute for Business Value. May 2011. www.ibm.com/cmostudy; *"Leading Through Connections: Insights from the IBM Chief Executive Officer Study."* IBM Institute for Business Value. May 2011. www.ibm.com/ceostudy

4 Woody, Todd. "Automakers, Tech Companies Mining Electric Car Big Data to Plot Industry's Future." Forbes. June 18, 2012. http://www.forbes.com/sites/toddwoody/2012/06/18/automakers-tech-companies-mining-electric-car-big-data-to-plot-industrys-future/

5 Ibid.

6 Ibid.

7 Ibid.

8 IBM case study. *"McLEOD RUSSEL INDIA LIMITED: Eliminating downtime in the tea trade with IBM Informix."* September 9, 2011. http://www-01.ibm.com/software/success/cssdb.nsf/CS/JHUN-8LFLWH?OpenDocument&Site=dmmain&cty=en_us

9 IBM case study. *"Premier Healthcare Alliance."* April 30, 2012. http://www-01.ibm.com/software/success/cssdb.nsf/CS/JHUD-8TS39R?OpenDocument&Site=wp&cty=en_us

10 IBM press release. *"Using IBM Analytics, Santam Saves $2.4 Million in Fraudulent Claims."* May 9, 2012. http://www-03.ibm.com/press/us/en/pressrelease/37653.wss

11 Kiron, David, Rebecca Shockley, Nina Kruschwitz, Glenn Finch and Dr. Michael Haydock, *"Analytics: The widening divide: How companies are achieving competitive advantage through analytics"* IBM Institute for Business Value in collaboration with MIT Sloan Management Review. October 2010. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-analytics-widening-divide.html © 2011 Massachusetts Institute for Technology.

12 Pittman, David. *"Lords of the Data Storm: Vestas and IBM Win Big Data Award."* The Big Data Hub: Understanding big data for the enterprise. September 28, 2012. http://www.ibmbigdatahub.com/blog/lords-data-storm-vestas-and-ibm-win-big-data-award

13 IBM case study. *"Automercados Plaza's increases revenue by 30 percent with greater insight into operations."* July 15, 2011. http://www-01.ibm.com/software/success/cssdb.nsf/CS/JHUN-8JPJSK?OpenDocument&Site=default&cty=en_us

IBM Global Services
Route 100 Somers, NY 10589
U.S.A.

Produced in the United States of America
October 2012

GBE03519-USEN-00