September 30, 2011

Expand Your Digital Horizon With Big Data

by Brian Hopkins and Boris Evelson
for CIOs

FORRESTER® Making Leaders Successful Every Day

September 30, 2011

Expand Your Digital Horizon With Big Data

by Brian Hopkins and Boris Evelson
with Sharyn Leaver, Connie Moore, Alex Cullen, Mike Gilpin, and Mackenzie Cahill

EXECUTIVE SUMMARY

At extreme scale, traditional data management and business intelligence (BI) become impractical, and your business does not get what it demands — more insight to drive greater business performance. Big data helps firms work with extremes to deliver value from data cost-effectively. However CIOs must understand that this is not business as usual. In fact, big data will disrupt the data management landscape by changing fundamental notions about data governance and IT delivery. Take the time to understand big data as well as its implications and begin a balanced approach that considers more than just the technology hype.

TABLE OF CONTENTS

NOTES & RESOURCES

Forrester interviewed various big data subject-matter experts and user companies, including Accenture, Appistry, Attivio, Basho Technologies, Bill and Melinda Gates Foundation, Composite Software, Endeca Technologies, Hortonworks, IBM, Infosys, Karmasphere, LexisNexis HPCC, PricewaterhouseCoopers, Saffron Technology, Splunk, TCS, Teradata, the University of Ontario Health Sciences Department, and Wipro. We also surveyed 60 clients on their plans for or use of big data technologies.

Related Research Documents

"How Forrester Clients Are Using Big Data" September 20, 2011

"It's The Dawning Of The Age Of BI DBMS" May 27, 2011

"How Social CRM Benefits From Big Data" May 24, 2011

"Big Opportunities In Big Data" May 18, 2011

YOUR FIRM'S FUTURE DEPENDS ON EFFECTIVELY USING MORE DATA

The amount of data available is growing faster than our ability to deal with it, and more is coming. As firms seek value from digital and analog sensors, social media, financial systems, emails, surveys, and customer call centers — to name a few — new tidal waves swell on the horizon.[1] All of this data promises to open up new frontiers of business opportunity — if we can figure out how to turn it into insight. Business-as-usual BI and IT delivery approaches, however, will not generate the future results your business expects.

Innovators Turn More Data Into More Value

Amid this swirling vortex of information, a few innovators have successfully tapped more data to wow customers, drive growth, achieve market leadership, unlock new revenue streams, and make our planet a better place. For example:

- **Google makes search easy by creating a new technology to deal with web scale.** In order to become the market leader in search, Google needed a way to deal with the massive volumes of web content. It developed a high-performance computing technology to meet its needs, and we all remember how quickly it rose to dominate its market as a result. Other Web 2.0 firms like Facebook and Yahoo followed suit, establishing an open source project to make this technology broadly available.

- **A hospital saves babies' lives using massive streams of monitoring data.** The University of Ontario is sponsoring research to collect nearly 100 million data points per day from premature babies and analyze them in real time. As a result, changes in patient vitals are correlated with the probability of sickness, allowing early action. Ultimately lives are saved that would otherwise have been lost.

- **A credit card company retains customers by understanding social relationships.** Studies have found that if one person defects from a product or service, other customers with social connections to that person may also. To capitalize on this study, a financial services firm is mining point-of-sale transactions at a massive scale to identify these social relationships based on card usage patterns to conduct targeted retention campaigns. As a result, it is increasing revenue and profit.

- **A telco taps into the Facebook social groups to market friends-and-family plans.** MicroStrategy recently rolled out Gateway to Facebook, a cloud-based service that allows its clients — such as a major telco — to enrich client and prospect data with Facebook's Social Graph and increase their chances to cross-sell/upsell friends-and-family phone plans.

- **A public utility performs sophisticated analysis on smart grid data.** Smart grids are collecting enormous volumes of data on our energy infrastructure, but even calculating average frequency can be challenging using traditional techniques. Using the same open source technology

pioneered by Google and Yahoo, the Tennessee Valley Authority implemented a system to analyze 1.5 trillion smart grid data points. As a result, it now performs sophisticated analysis on power grid anomalies that improves efficiency and ultimately saves natural resources.

- **IBM's Watson ushers in a new era of human-computer interaction.** In 2011, Watson stunned the world by beating two champion Jeopardy players. The computer not only was able to produce probable answers from straightforward questions but also recognized puns and adapted to nuances of language never before seen in artificial intelligence. Watson accomplished this using a new data storage and analysis platform that enabled massive-scale data crunching. The potential benefits of natural language human-computer interaction are clear and wide — from the mundane, such as shopping at the supermarket, to assisting complex medical diagnosis.

Four Characteristics Make Extreme Scale Difficult

Despite the opportunity, firms effectively use only a small fraction of the data currently available to them — and even more data is becoming available every month. One super-major energy company reported using fewer than 5% of the potential 25,000 data points per second available from an operating oil rig. Other clients anecdotally report similar numbers. Why? As data reaches the extreme, traditional solutions become expensive to scale or adapt to rapidly changing conditions. The added cost and time erode business cases for technology investment resulting in unrealized value. Business-as-usual methods are insufficient when:

- **The volume exceeds what can be cost-effectively stored.** As volume expands, traditional data platforms must scale to deal with it. For example, today's data warehouse (DW) appliances can reach well into the petabyte scale, but they rapidly become expensive. In some cases, firms may want to store petabytes of data for exploration and discovery where only a notion of value exists — rarely can the DW be justified for this.

- **The velocity of change prohibits timely decisions.** For example, an application processing a stream of sensor data may enable responses to critical events within hours. If the need is to take action within minutes, it may not be met, as scaling up becomes too expensive. Similarly, traditional decision support applications are usually based on fixed schemas, which take time to change. When business requirements evolve rapidly, these systems cannot affordably adapt.

- **The variety of formats makes integration expensive.** Integration costs grow along with the number of data formats because rules must be developed for each, and integration processes must be changed and redeployed. The cost of adding new data feeds makes traditional data integration methods too expensive for some scenarios.

- **The variability of data structures produces results that are hard to interpret.** For example, natural language search requires interpretation of complex and highly variable grammar. In BI, highly variable data structures make analysis difficult. For example, manufactures and retailers have difficulty analyzing a diverse portfolio of products or parts because they may have thousands, each with a unique set of characteristics.

In each of these situations, the common theme is cost. And it is not that current mature technologies cannot meet the need; it's that they cannot be cost-justified based on the benefits and risk.

BIG DATA MAKES EXTREME SCALE ECONOMICAL

Enter big data, which CIOs can tap to deliver affordable high-performance computing for complex extreme-scale operations such as long-tail analysis, microsegmentation, next-best offer/action, web session identification, and customer experience optimization, to name a few.[2] Forrester defines big data as:

> *Techniques and technologies that make handling data at extreme scale affordable.*

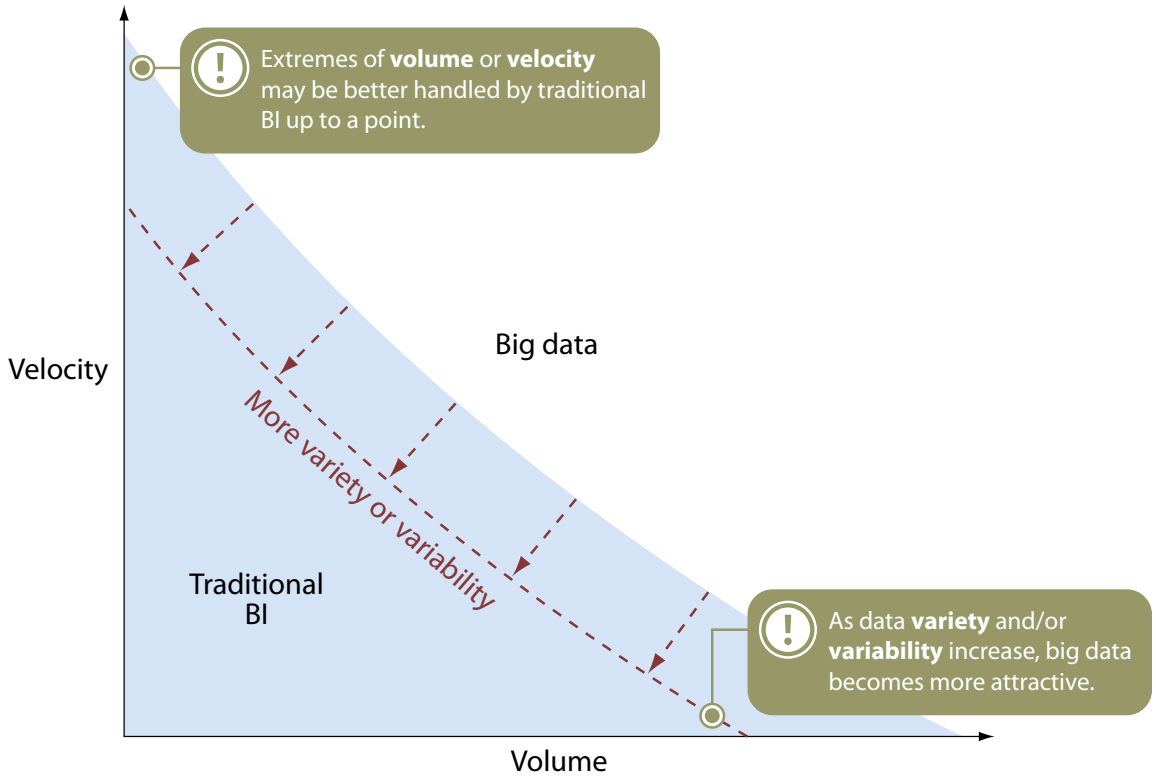
Despite its simplicity, this definition is not trivial. Forrester asserts that big data 1) is not only technology but also people with the appropriate analysis skills, and 2) makes dealing with extreme scale affordable. By making it cost-effective, more firms can adopt big data to increase their competitive advantage, which leads to more innovation, more vendor solutions, and more momentum, among other benefits.

Big Data Tackles The Four V's: Volume, Velocity, Variety, And Variability

Contrary to common market misconceptions, big data is not just about volume. Forrester recently surveyed 60 clients with some knowledge and/or experience with big data.[3] While three-quarters of the respondents indicated extreme volume as the main reason for considering big data solutions, others indicated the velocity of change, format variety, and structural variability as major concerns. One of these characteristics alone may justify pursuit of big data, but if several exist together, the technology becomes increasingly attractive (see Figure 1).

Figure 1 The Four V's Determine When Big Data Should Be Considered

Volume — approaches or exceeds physical limits of vertical scalability	**Velocity** — decision window small, compared with data and/or data requirements change rate
Variety — many different formats that make integration expensive	**Variability** — many options or variable interpretations that confound analysis



60751 Source: Forrester Research, Inc.

Big Data Technology Is All About Discovery

Firms are developing new Agile BI practices using big data. For example, 58% of respondents in Forrester's June 2011 Global Big Data Online Survey reported interest for "analysis-driven requirements discovery" reflecting a "*build it and they will come*" approach. Analysis of big data leads to new insights, which are used to establish requirements for more traditional BI solutions. This approach is backward from typical approaches, but we find many firms adopting it because:

- **Data capture and storage is easy and cheap.** Moving data into a big data store is often a one-step process that can be set up in less than a day. For example, Yahoo uses open source tools to capture in 3 hours about 450 terabytes of new data every day. New data feeds can be added in about 30 minutes with a few lines of script. While this data is not aggregated, cleaned, or integrated, big data techniques enable teams of data scientists to conduct sophisticated analysis and deliver value rapidly.

- **Analytics get to insight quickly by massive parallel processing.** Stored data can be analyzed by parallel processing frameworks. For example, data scientists can experiment with mining techniques by lifting and transforming petabytes of data as easily as writing a few lines of script. Sophisticated tools on top of parallel computation frameworks are empowering business experts to rapidly build analytic applications.

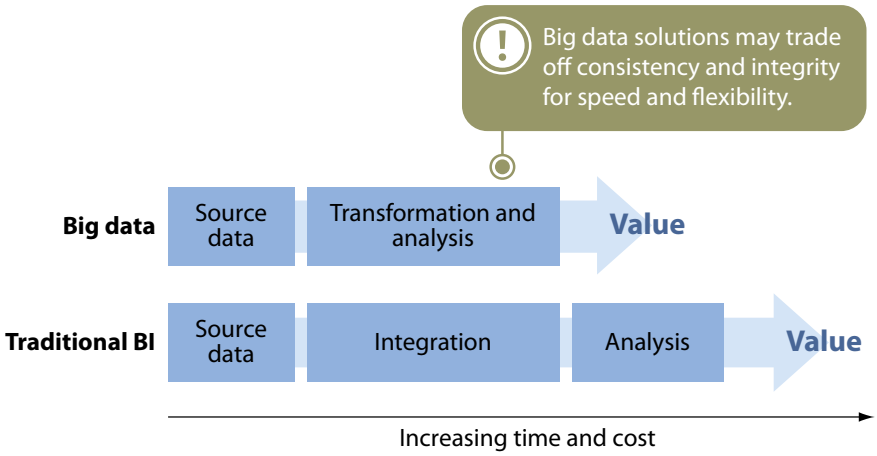
- **Platforms are highly scalable using commodity infrastructure and cloud.** Big data technology is built to operate reliably across vast arrays of commodity computing resources despite the inevitable failures. For example, Yahoo's big data implementation consists of 100,000 CPUs on 40,000 commodity computers. Many others boast a few hundred to a few thousand. Because big data technology generally works on commodity, many firms are electing to use public infrastructure-as-a-service (IaaS) rather than buy the hardware. This further enhances the business case by avoiding capital expenditure to get up and going with big data.

Big Data Gets To Insight Faster When "Two Plus Two Equals 3.9" Is Good Enough

Big data is far different from traditional BI solutions (see Figure 2). In the latter, extract-transform-load (ETL) operations are performed to clean, conform, and integrate multiple data sources before analysis is performed. This use case remains important when exact certainty is required — when two plus two must equal four — and it's always a case in, say, month-end closing calculations, even if it takes weeks to process and millions to build and maintain. But other situations do not require the same rigor. Sometimes two plus two can equal 3.9, and that is good enough. To a marketing executive who wakes up in the morning to a new competitive threat and needs to immediately counter it with his own campaign, speed of customer segmentation for the campaign trumps accuracy. For these types of analysis, big data solutions can often deliver the approximate answer much faster.

Because of its advantages, we found that 70% of respondents are interested in big data for handling existing enterprise information. This tells us that while the hype is about complex analysis of new data sources, the truth is that many early adopters are taking a pragmatic approach, using it where it makes sense with their existing information.

Figure 2 Big Data Often Operates On Raw Sources To Deliver Value More Quickly



60751 Source: Forrester Research, Inc.

BIG DATA WILL GET EASIER AS PEOPLE, PROCESSES, AND TECHNOLOGY MATURE

Our research confirmed that big data, in its current immature state, is challenging. Successful firms will meet this challenge through 1) establishing a new level of collaboration between business and IT; 2) developing new processes to deliver solutions; and 3) mastering a rapidly evolving technology landscape.

Business Leaders, Data Scientists, And IT Will Collaborate On Big Data Initiatives

In Forrester's June 2011 Global Big Data Online Survey, 70% of respondents said that big data is or will be a collaborative effort between business and IT. Our interviews revealed big data teams of highly skilled scientists working closely with business executives to discover new uses for more data. In order to succeed in this brave new world of big data, CIOs must:

- **Partner with their business peers to identify opportunities and solutions.** While business executives at some firms immediately get it, we expect the concepts of big data to be foreign to most. CIOs who understand the state of the art and are perceived as strategic partners will be well positioned to help executives justify investments and reap new benefits. Conversely, CIOs without a seat at the business strategy table may struggle, as big data solutions are not business-as-usual BI.

- **Turn over part or all of big data solution delivery to business leaders.** A recurring theme in our research was the level of business involvement in big data solutions. One interviewee recounted big data environments owned by other departments, such as marketing, with no IT involvement. In most cases, IT has a role to ensure that big data technology meets service-level agreements. However, firms with a cadre of technology-savvy business people may want the power to create their own analytic solutions.

- **Be ready to support rapid growth.** A second recurring theme in our research was the speed at which big data solutions can take off. Interviewees and case studies we reviewed told similar stories — firms began a big data solution as a pilot with a few terabytes of planned growth, only to find themselves with more than a petabyte very quickly. Since the same data can be used different ways and reanalyzed for new insights easily, nothing ever gets deleted. In the words of one shipping industry executive: ". . . No data is discarded anymore . . . [we] leverage a large scale of transaction data and a diversity of interaction data . . . [and] saved millions of dollars per year by . . . augmenting our enterprise data with sensor, meter, RFID tags, and geospatial data."

New Processes Will Deliver Tomorrow's Big Data Solutions

The unique characteristic of big data will drive new processes. We found that nearly one-third of firms surveyed are currently using different project management and software delivery life cycle (SDLC) processes, while nearly as many are not sure what they will do once efforts move beyond prototyping. For example, new big data initiatives will require:

- **New processes to manage open source risks.** Many big data solutions are being built on open source software. But open source has both legal and skill implications as firms are 1) exposed to risk due to intellectual property issues and complex licensing agreements; 2) concerned about liability if systems built on open source fail; and 3) required to use technology that is often early-release and not enterprise-class. This is especially high concern for firms in highly regulated industries such as financial services and pharma.

- **New Agile processes for solution delivery.** Big data solutions are analysis-driven and must be flexible to meet rapidly evolving requirements. Typical requirements-led solution delivery via waterfall SDLC practices will prove to be too cumbersome. Successful firms will embrace Agile practices that allow end users of big data solutions to provide highly interactive inputs throughout the implementation process. The more end users can control the outcome of big data solutions, the better they will be received.

- **New security and compliance procedures to protect extreme-scale data.** Existing row, column, and field-based data security models as well as compliance support tools such as data encryption and masking are based on traditional database designs. Since big data is a completely new architecture, your tools and the processes that use them will not work. In order to succeed with big data, new processes must be developed that recognize and protect the special nature of extreme-scale data that may be largely unexplored.

Open Source Leads The Technology Market, But Other Options Are Expanding

The name big data originated as a tag for a class of technology with roots in high-performance computing, as pioneered by Google in the early 2000s. Today, the big data market is expanding quickly and includes:

- **Distributed file and database management tools led by the Apache Hadoop project.** The current big data market centers on the Apache Hadoop open source initiative, which provides a distributed file management system that is the foundation of most big data solutions. Other alternatives to Apache exist, and more are emerging — for example, Aster Data and Microsoft High Performance Computing (HPC).

- **Big data analytic platforms, also led by Apache.** In addition to file management, Hadoop provides a computational framework for executing massively parallel analytic jobs on data in the file system. As with file management, other vendors have created similar analytic frameworks, such as Appistry's CloudIQ and LexisNexis High Performance Cluster Computing (HPCC).

- **Integration technology for exposing data to other systems and services.** Once big data has been captured and analyzed, firms must integrate the results back into their other information systems. Vendors like Informatica have incorporated big data into their tools using open source integration capabilities like Apache Hive, which provides structured query language (SQL) access to big data.

- **DW and BI tools with integrated big data capability.** BI and DW vendors have jumped into the big data fray by including Hadoop distributions with their packages, making it easy for users to capture and analyze big data. For example, Jaspersoft and Pentaho provide a GUI-driven interface for designing analytic jobs that would require coding if the native Apache distribution was used instead.

- **Stream processing and analytics tools.** Forrester's definition of big data does not limit it to batch-oriented store and process technologies like Hadoop. Stream analytic solutions such as Apache S4 and IBM Infosphere Streams do not store the data. Rather, they support high-velocity use cases by analyzing the data as it flows through a system. These technologies, like their batch-oriented counterparts, leverage massive parallel processing and commodity infrastructure to achieve operations at extreme scale.

Big data technology, while early-stage, is not vapor-ware, and various vendors offer production-ready solutions (see Figure 3).

Figure 3 Examples Of Big Data Vendors And The Capabilities They Provide

Big data technology	Example vendors
Big data file and database management	Appistry, Aster Data, Basho, Cassandra, Google, Hadoop (Hortonworks, Cloudera, MapR), LexisNexis, Microsoft, VMware
Big data analytics	Appistry, Aster Data, Hadoop (Hortonworks, Cloudera, MapR), IBM, LexisNexis, Karmasphere
Big data integration	Appistry, Composite, Datameer, Hadoop Pig and Hive, IBM, Informatica, LexisNexis
DW appliances with big data integration	EMC/Greenplum, IBM/Netezza, Teradata/Aster Data
Traditional BI with big data integration capability	Endeca, Jaspersoft, Pentaho, MicroStrategy, Tableau Software
Stream processing and analysis	Apache S4, IBM, Progress, SAP, StreamBase, Tibco Software

60751 Source: Forrester Research, Inc.

Ease Of Use, New Analysis, And Real Time Are The Next Frontiers

Today, big data technology can be daunting, requiring open source tools, lots of arcane command line configuration, and coding. This is not uncommon for early-stage disruptive technologies — we all remember DOS, which eventually gave way to Windows. As early-adopters' success leads to the early majority investment, vendors are rushing to make big data technology both easier and more powerful by:

- **Providing ways to deliver solutions other than writing complex code.** Vendors are rushing to overcome technology complexity by 1) building tools that access big data through familiar query languages like SQL; 2) providing graphical design layers to minimize or eliminate coding; and 3) adding big data functions to tools that users already have. For example, DW appliance vendors such as EMC/Greenplum are building in access to big data stored in Hadoop so that clients can move insights from big data analytics into their DW.

- **Supporting different types of analysis.** Today, many big data solutions employ a simple computation model know as MapReduce. However, not all types of analysis are suitable for it, greatly increasing the amount of effort required to implement solutions.[4] We found that new models appear in tools that greatly simplify complex operations, which MapReduce does not handle well. As an example, the Apache Giraph project is developing a social graph analysis capability based on Hadoop, which will make it much easier to extract insight from complicated social relationships for customer marketing and retention campaigns.[5]

- **Moving from batch to real time.** Most big data technology is batch-oriented and currently cannot meet requirements for interactive analytics and real time. However, the technology is evolving to overcome these limitations. As a result, future big data solutions will easily operate as part of production business-critical applications, offering new opportunities for competitive advantage. As an example, one solution we identified uses big data to perform genetic analysis as part of a medical diagnosis, reducing the time required from weeks to minutes. This capability allows doctors to diagnose conditions and recommend treatments in near-real time.

FRAME BUSINESS OPPORTUNITIES USING THESE FIVE SCENARIOS

Big data requires close collaboration between business and IT, as neither has the full picture. The focus of this collaboration is to match potential opportunities from the business perspective with the approach for realizing these opportunities. We found five common usage patterns that describe how firms generally use big data technologies (see Figure 4). Keep these in mind as you hold business conversations.

Figure 4 Common Patterns For Employing Big Data

Pattern	Description	Examples
Exploration and machine learning	This involves iterating on large data sets and looking for patterns and new ways to predict future outcomes.	Social relationship analysis for churn prevention, market microsegmentation analysis, multivariant risk pricing
Operational prediction	Big data feeds operational predictive models with new data upon which to base predictions.	Search result ranking, medical diagnosis based on genetic profiling
Dirty ODS	Raw data, of limited consistency and cleanliness, is accessed for operational reporting where "good enough" is good enough.	Standard reports generated from customer call records for operational purposes
Bulk data operations and extreme ETL	Batch operations on data at massive scale are conducted using parallel processing techniques.	Creating PDF files from terabytes of raw images, making data warehouse operations faster and cheaper with massive-scale bulk data movements
Stream and event analytics	Rapidly changing data is processed in parallel using complex events or more sophisticated stream filtering and mining algorithms.	Equipment monitoring and failure analysis, medical monitoring, logistics analysis from RFID tags.

60751 Source: Forrester Research, Inc.

RECOMMENDATIONS

START SMALL — WITH AN EYE TO SCALE QUICKLY

While big data technology may be quite advanced, everything else surrounding it — best practices, methodologies, org structures, etc. — is nascent. No one has the answers — not yet. However, Forrester recently published some research for you to see what those questions are and how you compare with other Forrester clients addressing big data challenges.[6] Additionally you must:

- **Understand why traditional BI/DW can't solve a problem.** Big data is not necessarily the only answer. If scaling your data servers up and out will solve a volume problem, it may be your best option. Furthermore, those in-memory analytic features of your BI tools or your DW appliance may be just the ticket for real-time and schema-less analysis. Use Forrester's four V's of big data and our five usage patterns to make sure that you are looking at the right opportunities.
- **Develop a minimal set of big data governance directives upfront.** Big data governance is a chicken-and-egg problem — you can't govern or secure what you haven't explored. However, exploring vast data sets without governance and security introduces risk. Address this by starting with directives to ensure 1) different groups or individuals don't analyze the same data to get different answers to the same question; 2) access is controlled to big data environments according to any protected information it may contain; and 3) compliance policies explain your firm's rationale for handling big data the way you intend to.
- **Remember the three S's — small, simple, and scalable.** When launching big data initiatives, avoid 1) getting too complicated too fast, and 2) not being prepared to scale once a solution catches on. Big data solutions can quickly grow out of control since discovering value from data prompts wanting more data. Keeping big data as small as reasonable, reducing scope to the simplest, most valuable objectives, and being ready to scale are secrets to success.

WHAT IT MEANS

CONSISTENCY AND CONTROL GIVE WAY TO CONSTANT CHANGE AND RELATIVE TRUTHS

Big data can drive big changes in your business model as you discover new ways to capture more value from more data. However, achieving success will also mean that:

- **A single version of the truth may not exist for all of your data.** Many firms depend on having central databases at the heart of their applications to provide a single, consistent version of the truth — the truth is that this architecture will not serve them as scales reach the extreme. Instead, we see designs emerging that trade off immediate consistency and integrity for scalability, availability, and fault tolerance using technologies such as big data and NoSQL. For these firms, the ability to operate at mega-scale outweighs other needs, and they are learning how to deal with local and relative truth.

- **Big data will redefine information governance.** The scale and nature of big data operations may render some traditional governance techniques inappropriate, exposing you to risk. First is the risk of making business decisions based on an incomplete or inconsistent view of relevant information. Second, compliance monitoring will be more difficult. Finally, you may not have visibility into unethical activity, criminal behavior, policy violations, quality issues, or other problems with big data. These are not show stoppers, but governance must clearly evolve to deal with a big data world.
- **Untamed processes will no longer be the enemy.** Firms have been ill-equipped to deal with vast, sprawling, and constantly changing business processes. Efforts to standardize and simplify often fail due to the business limitations imposed. Big data provides tools to understand complex relationships and make decisions despite lack of perfect knowledge — invaluable for dealing with process complexity. In Forrester's June 2011 Global Big Data Online Survey, only 16% of respondents indicated that business process management and business rule technologies are in the scope of their big data projects. We believe that this will change as firms discover that big data helps them deal with processes that cannot be tamed.

SUPPLEMENTAL MATERIAL

Methodology

Forrester's June 2011 Global Big Data Online Survey was fielded to 60 IT professionals. Forrester fielded the survey during June 2011.

Exact sample sizes are provided in this report on a question-by-question basis. Unless otherwise noted, statistical data is intended to be used for descriptive and not inferential purposes.

Companies Interviewed For This Document

Accenture

Appistry

Attivio

Basho Technologies

Bill and Melinda Gates Foundation

Composite Software

Endeca Technologies

Hortonworks

IBM

Infosys

Karmasphere

LexisNexis HPCC

PricewaterhouseCoopers

Saffron Technology

Splunk

TCS

Teradata

University of Ontario Health Sciences Department

Wipro

ENDNOTES

[1] According to the EMC-sponsored annual "2011 Digital Universe Study: Extracting Value from Chaos," the amount of information created in 2011 will be 50% more than in 2010, with a projection that information production will double approximately every two years. According to a study sponsored by Oracle, nearly half of companies surveyed experienced less than 25% data growth during 2010. Even though these numbers are only estimates, the discrepancy is clear. Data is being digitized nearly twice as fast as firms are able to capture it in their information systems. Even as of 2009, mobile devices were generating 600 billion geospacially tagged transactions per day. In 2011, Twitter reached 140 million tweets per day. Even if the number of mobile devices in the world has stayed relatively constant, that is still more than 4,000 times more geolocation data points than tweets. Source: John Gantz and David Reinsel, "2011 Digital Universe Study: Extracting Value from Chaos," EMC, June 28, 2011; Joseph McKendrick, "Keeping Up With The Ever Expanding Enterprise Data," Oracle, October 2010; ReadWriteWeb (http://m.readwriteweb.com/archives/meet_the_firehose_seven_thousand_times_bigger_than.php).

[2] Long-tail analysis refers to the retail strategy of selling a large number of unique items in relatively small quantities; customer microsegmentation involves creating highly detailed customer segments and is usually done as part of long-tail analysis; next-best offer/action uses predictive models to automate customer service actions as part of CRM; session identification is a process of linking and correlating multiple user sessions — clicks — on an eCommerce website to identify a complete end-to-end path a user took to buy a product or abandon a shopping cart.

[3] Forrester has published data from the June 2011 Global Big Data Online Survey in a companion report. See the September 20, 2011, "How Forrester Clients Are Using Big Data" report.

[4] MapReduce: It is a two-step process in which find, filter, and sort-type operations are distributed, or "mapped out," among many thousands of jobs running on many computers at once. Then the results are "reduced" to a single answer by a master system. This operation is just one of many high-performance computing algorithms possible, and it is not well suited to all types of analysis.

[5] Social graph analysis is used to mine data from social sites like Facebook to identify relationships for marketing and customer relationship management (CRM).

[6] As a supplement to this report, Forrester reviewed and analyzed the results of the June 2011 Global Big Data Online Survey. See the September 20, 2011, "How Forrester Clients Are Using Big Data" report.

FORRESTER®

Making Leaders Successful Every Day

Headquarters

Forrester Research, Inc.
60 Acorn Park Drive
Cambridge, MA 02140 USA
Tel: +1 617.613.6000
Fax: +1 617.613.5000
Email: forrester@forrester.com
Nasdaq symbol: FORR
www.forrester.com

Research and Sales Offices

Forrester has research centers and sales offices in more than 27 cities internationally, including Amsterdam, Netherlands; Beijing, China; Cambridge, Mass.; Dallas, Texas; Dubai, United Arab Emirates; Frankfurt, Germany; London, UK; New Delhi, India; San Francisco, Calif.; Sydney, Australia; Tel Aviv, Israel; and Toronto, Canada.

For the location of the Forrester office nearest you, please visit: www.forrester.com/locations.

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com.
We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 19 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 28 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit www.forrester.com.

FORRESTER®