CHAPTER 1.9

# Making Big Data Something More than the "Next Big Thing"

**ANANT GUPTA**
HCL Technologies

*Big data is the business buzzword du jour. But how can you turn this hot topic into a real source of business value?*

You have certainly heard the breathless spiel: The world today is being flooded with digital data, in myriad manifestations and washing over us at such incredible speed that making sense of it is dauntingly difficult. Yet this tidal wave of data—when channeled and filtered by an array of new information technologies—holds untold value for organizations, whether they are small not-for-profits or Fortune 500 companies.

Or so we are told. But despite the sometimes exaggerated hype surrounding "big data," the fundamental assertion is true: data—and the decisions driven by those data—now represent the next frontier of innovation and productivity.

Estimates of the potential benefits of leveraging big data are indeed staggering: productivity-led savings worth US$300 billion a year for the US healthcare industry and €250 billion for the European public sector, a 60 percent potential increase in retailers' operating margins.[1] And technology seems poised to deliver these benefits. One small example: data storage technology has advanced to the point that only US$600 is all it takes to purchase storage space that can accommodate the entire world's music![2]

Some large companies have indeed used emerging technologies to extract significant value from big data. Visa recently announced that increasing from 40 to 200 the number of attributes it analyzes in each credit card transaction has saved 6 cents in every $100 worth of transactions.[3] Wal-Mart uses a self-teaching semantic search tool that, honed by the monthly clickstream data of 45 million online shoppers, tailors offerings to online shoppers, raising the rate of completed transactions by more than 10 percent.[4]

But for most businesses, the promise of big data is nowhere close to being fulfilled. For one thing, spending on it is polarized. While the telecommunications, travel, retail, life sciences, and financial services industries are making significant strides in big data technologies, other industries, such as manufacturing and government,[5] are in a wait-and-watch mode.

The lack of major big data initiatives across industries can be seen in the numbers from service providers. In 2012, the global top 20 big data players made *less than 1 percent* of their total revenues from big data. The total market for big data hardware, software, and services in 2012 was US$11.5 billion, whereas the combined overall revenue of those 20 big data players was more than US$1.2 trillion.

The disparity between a few success stories and the lack of action elsewhere has created a high level of anxiety within firms that have not yet begun to explore big data. But it is important that they not rush thoughtlessly into the fray. An organization should make

© 2014 World Economic Forum

### Box 1: A user's glossary of key big data terms

As an organization plans its big data strategy, the following terms are likely to be used with increasing frequency.

- **Hadoop:** A batch-oriented programming framework that supports the processing of large data sets in a distributed computing environment. Hadoop is written in the Java programming language and is a top-level Apache project (Apache is a decentralized community of developers supporting open-source software).
- **HBase:** A non-relational, column-oriented distributed database written in Java. A column-oriented database stores data tables as sections of columns of data rather than as rows of data, as in most relational databases, providing fast aggregation and computation of large numbers of similar data items.
- **HDFS:** A distributed, scalable, and portable file system written in Java for the Hadoop framework.
- **Hive:** A data warehouse infrastructure built on top of Hadoop, providing data summarization, query, and analysis. It permits queries over the data using a familiar SQL-like syntax.
- **Flume:** A tool for collecting, aggregating, and moving large amounts of log data from applications to Hadoop.
- **Mahout**: A library of Hadoop implementations of common analytical computations.
- **Oozie:** A workflow scheduler system developed to manage Hadoop jobs.
- **Pig:** A platform for analyzing large datasets that consists of a high-level language (Pig Latin) for expressing data analysis programs, coupled with infrastructure for evaluating these programs.
- **R:** R is a free software programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.
- **Sqoop:** A tool facilitating the transfer of data from relational databases into Hadoop.
- **Zookeeper:** A centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services for distributed applications.

**Source:** HCL 2013b.

a big data investment only if it has well-defined and realizable business objectives.

We offer here nine steps that companies can take to begin turning big data talk into action, buzz into business benefits.

## WHY IS EXTRACTING VALUE FROM BIG DATA SO HARD?

First, though, we examine some of the barriers to realizing big data's promise.

### Volume, velocity, and variety

Big data is often said to be characterized by 3 Vs: its tremendous volume, the velocity at which it needs to be processed, and the variety of data types it encompasses. The first two characteristics are fairly obvious: technology has made it possible to capture increasingly large amounts of information and make it available for analysis in real time.

But mining the value of big data also is difficult because it requires simultaneously analyzing various types of information—transactions, log data, mail documents, social media interactions, machine data, geospatial data, video and audio data, to name just a few—much of which is "unstructured." Traditional types of business data were available in a format that was structured and could have been automatically analyzed—for example, a spreadsheet quantifying customer returns of different products at different stores over time. However, much of the value in big data exists in unstructured information—for example, the transcript of a chat session between a retail customer and a customer service representative.

Synthesizing unstructured data from numerous sources and extracting relevant information from it can be as much art as science.

### Talent scarcity

Much has been said and published about the looming talent gap. Estimates suggest that the United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million analysts and managers to analyze big data and make decisions based on those findings.[6] Another report predicts that only one-third of 4.4 million big data jobs created by 2015 will be filled.[7] Unlike traditional analytics, mining big data requires an extremely diverse set of skills—deep business insights, data visualization, statistics, machine learning, and computer programming. Policy should work to mitigate this talent shortage through forward-looking education and immigration policies.

### Flawed data governance

Big data is not a substitute for—much less a solution for—flawed information management practices. If anything, it requires much more rigorous data governance structures. Without those improvements, information technology (IT) systems that have not been upgraded to handle large volumes of data are likely to collapse under the sheer weight of the data being processed. Surveys suggest that business leaders are often more excited about the potential of big data


© 2014 World Economic Forum

than their IT counterparts. That may be because of IT executives' understanding of the realities on the ground.

### Lack of a data-driven mind-set

Because mind-set can be hard to pin down, its power is often underestimated. That is a mistake when it comes to assessing the prerequisites to successful analytics deployment. It is virtually impossible for big data investments to deliver value if business leaders do not have a data-driven mind-set—that is, if they do not believe that it is important for decisions to be based on cold, hard numbers rather than gut feel and experience. But once the right mind-set takes hold, other good things will follow: data-driven business leaders will have a tremendous incentive to treat data, and therefore the IT and analytics professionals who help deliver it in an understandable form, as a strategic asset. And these leaders will make it a priority to ease the flow of data across organizational silos.

### Lack of technical know-how

Big data represents a convergence of IT and data science. Technologies include Hadoop (which enables large-scale processing of diverse datasets), R (a programming language for statistics), and in-memory databases (where data reside on main memory as opposed to disk storage). Data science includes, among many other areas, machine learning (systems that learn from data) and data warehousing. Big data professionals are expected to be familiar with both disciplines, but this combination is rare, despite the training courses that are sprouting up globally. (For descriptions of some of the technologies that enable the analysis of big data, see Box 1.)

## NINE STEPS TO BIG DATA VALUE CREATION

The barriers to extracting business value from big data can seem daunting. But they can be overcome through a systematic plan, one that breaks down the challenge into a series of nine sequential steps that will enable organizations to take advantage of this valuable and growing asset. We will consider each of these steps individually here.

### Step 1: Define responsibilities.

Who collects, who analyzes, and who drives value? The onus of collecting data should be shared by the IT and analytics teams, but analysis must be the sole responsibility of analytics professionals. Similarly, only functional leaders—for example, the Chief Marketing Officer, the Chief Financial Officer, and the Chief Procurement Officer—should be responsible for identifying areas within their respective functions where big data could drive value. However, getting this level of support from functional leaders is not easy, especially if the team—IT and analytics or a dedicated big data center of excellence—reside outside of the business function. In order to drive the big data program, the team may want to appoint a big data program sponsor for each function and work closely with him or her to discover and locate the types of information that would improve business outcomes. Most importantly, however, the program sponsor would try to get functional buy-in and identify big data opportunities within the function.

### Step 2: Get the business functions to ask the right questions.

Senior executives will have an easier time winning buy-in from business functions if they demonstrate how big data might be valuable to them. Simple questions such as "What would you really like to know about your business, and how can data help you with it?" are a good place to start. Such questions can spur the functional experts themselves to start asking the more fundamental questions that can unlock the value of data. For instance, marketing professionals could ask, "What is the value of a 'tweet' or a 'like'? Are our investments in customer service paying off? What is the optimal price for our product right now?" The ability to ask the right questions is key to succeeding with big data. It also pays to keep in mind that big data is *not* about data themselves; it is about using data to discover insights that can lead to valuable outcomes.

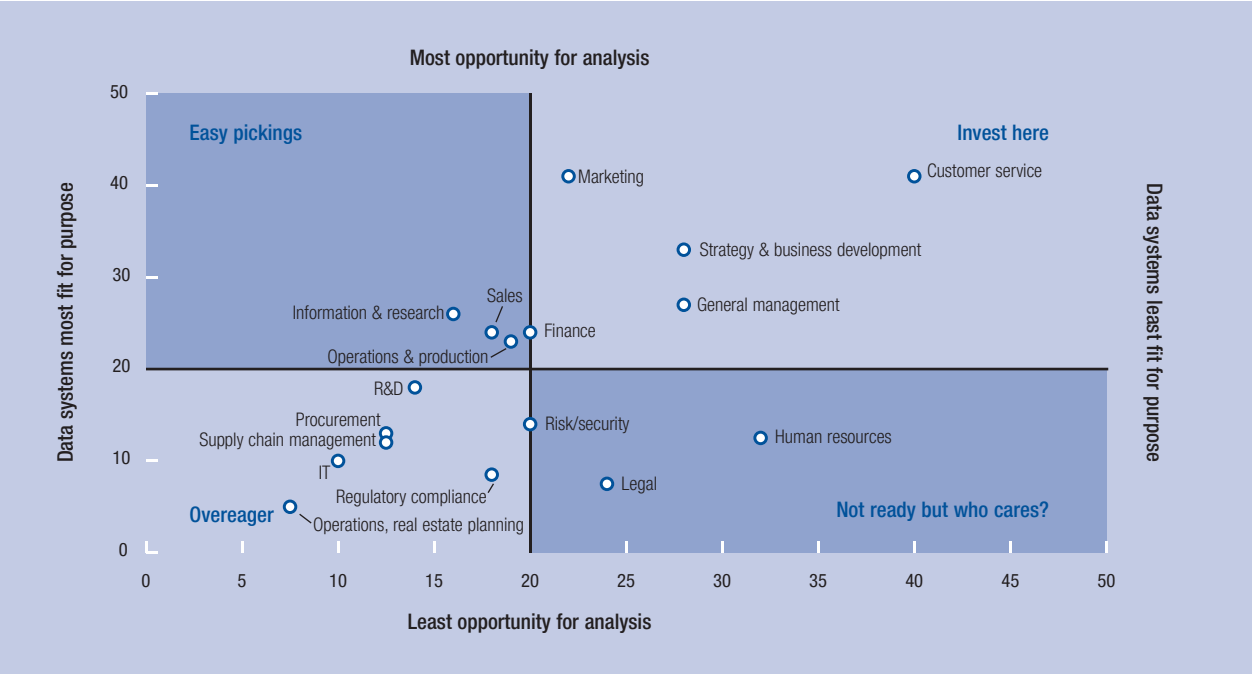
### Step 3: Take stock of all data "worth analyzing."

Valuable business insight can come from many sources, including social media feeds, activity streams, and "dark data" (data that are currently unused but that have already been captured), machine instrumentation, and operational technology feeds. It is important to explore these sources and to experiment with new ways of capturing information, such as complex-event processing, video search, and text analytics. Organizations' data typically fit into four buckets:

- Operational data, such as data emanating from smart grid meters, embedded systems (examples include microwave sensors and chips inserted in automobiles), transactions logs (such as payment transactions), radio-frequency identification chips (RFID), navigation and location sensors, networks, and servers.
- Streaming data, such as computer network data, phone conversations, and so on.
- Documents and content, such as PDFs, web content, and legal discovery elements (electronic information exchange in civil litigations).
- Rich media, including audio and video tracks, electronic images, and so on.

© 2014 World Economic Forum

**Figure 1: Potential payback of big data initiatives**




Source: Gartner, 2013.

### Step 4: Select the business functions best positioned to lead the way.

It is smart to launch big data initiatives in business functions that are most ready to collect and analyze data and for which the potential payback is high. Functions such as marketing, customer service, supply chain management, and finance are poised for maximum growth. If system readiness is not an issue, these are usually the right places to direct initial investments (see Figure 1).

### Step 5: Match big data initiatives with compatible business functions.

Some big data programs can be implemented in a variety of settings, but most are suited to specific functions. For example:

- Customer functions (such as marketing, e-commerce, and customer service) can use big data for targeted advertising that provides personalized offers to consumers based on their socio-demographic characteristics, and for loyalty management that extends channel reach from point of sale, web, and call center to include mobile and social capabilities.
- Finance functions (such as finance, risk, and treasury) can use big data for intraday liquidity management, providing real-time monitoring of price movements in relation to positions, to make trading and rebalancing decisions, and for improved credit risk assessment, through multiple big data–supported credit risk assessments that factor in hundreds or even thousands of indicators.
- Supply chain and procurement can use big data for dynamic route optimization because big data technologies that are faster than conventional systems allow more iterations and faster route planning in real-time.

### Step 6: Determine whether big data will yield valuable information unavailable through traditional business analytics.

Making the business case for a big data initiative clearly will be easier if it can be shown that it creates new value. For instance, if a marketing department is currently segmenting customer profiles using standard demographic indicators, would there be additional benefit in analyzing attitudes and preferences (at a granular level) through text and speech analysis? Similarly, if a traditional business intelligence program is currently analyzing financial market sentiments using structured stock information, would it make the sentiment analysis more refined by including social media feeds, news sites, and so on?

In comparing views of data from a traditional business intelligence perspective versus a big data one, consider the following the questions: What data are we capturing today? What are the limitations of this kind of structured data? What extra value will we get by collecting external, context-specific, and unstructured data? Where will we find data and how will we collect them? Would our business act upon the insights


© 2014 World Economic Forum

gained? Is the extra business value worth the additional investment of time, energy, and money?

### Step 7: Assess complexities and prioritize accordingly.

All else being equal, an organization should begin its big data experimentation with an initiative that is not too demanding. In assessing possibilities, it is helpful to keep in mind the complexity of both the *type of data* and the *type of analysis* the data will require.

As we mentioned above, much of what is meant by "big data" is unstructured information—data that traditionally have been impossible to break down and categorize as they are collected. Such data are not only difficult to analyze but can also be easily misinterpreted when taken out of context. Thus it makes sense to experiment in the beginning with data that are relatively easy to analyze.

Different types of analysis also present varying degrees of complexity. Generally speaking, descriptive analytics (which answers "what happened?"—for example, an analysis of social media sentiment analysis) are relatively easy to do. However, diagnostic analytics (which answers "why did it happen?"—for example, an analysis of customer defection at the shopping cart stage of the online purchase process); predictive analytics (which answers "what will happen?"—for example, forecasts of customer churn in telecommunications); and prescriptive analytics (which answers "how can we make it happen?"—for example, determining whether personalized offers to customers would make sense), are increasingly complex to conduct.

### Step 8: Assess your technology architecture.

An organization's traditional information architecture may not accommodate massive, high-speed, variable data flows. Many traditional and even state-of-the-art technologies were not designed for today's or tomorrow's level of data volume, velocity, and variety. Even as datasets grow exponentially along those dimensions, the investments required for scaling technologies (such as processors, storage, database management systems, and analytics) to perform efficiently grow even faster. To counter these intractable economics, organizations need to consider a variety of methods to upgrade their infrastructure in support of or in anticipation of big data.

In fact, the idea that big data involves negligible cost because it is analyzed using open-source tools and platforms is a myth. "Free" open-source technologies such as Hadoop (which enables large-scale processing of diverse datasets) are typically not immediately usable. You need either to hire and train data scientists and analysts in Hadoop programming, or to buy an enterprise-ready version of Hadoop.

If the outcome of big data analysis is mission-critical for your business, it probably makes sense to use only purpose-built hardware. Generic servers may be fine for smaller projects and proofs of concept, but specifically designed, enterprise-grade servers, storage, and networking products are best for large-scale-production solutions.

Creating or upgrading to big data–ready technology architecture is no small feat. Building everything from scratch takes time, and buying everything is expensive. Therefore, finding the right combination of insourcing and outsourcing requires careful consideration.

### Step 9: Start building a team.

Big data initiatives require multidisciplinary teams of business and technology experts. Every team member—business analyst, programmer, data scientist, and data visualizer—will need to have cross-functional familiarity. Building this team is a five-step process:

- Break down your talent needs into four distinct areas: business analysis, analytics, database technology, and data visualization.
- Scan your internal landscape for the aforementioned skills. Although they may not be in the target department, every organization probably already includes people who know the business, possess data-crunching capabilities, and make data-driven decisions.
- Hire people with needed skills if they are not available or cannot be acquired by cross-training existing employees.
- Hire people with related skills if the needed skills are unavailable within your organization or difficult to acquire through external hires. For instance, consider substituting statisticians for the much less common data scientists.
- Start small and scale up. In the beginning, your needs will be modest. A few hires may be adequate to get started.

Some are even predicting that big data analytics will lead to the emergence of an entirely new set of CXO roles within enterprises—Chief Data Officer, Chief Digital Officer, Chief Analytics Officer, and so on. That said, the structure of most organizations would make it difficult for someone owning the big data portfolio to succeed. Without clear line responsibilities, a CDO (whichever flavor, Data or Digital) or a CAO would have little leverage to execute the important tasks needed to increase the organization's big data capabilities and optimize its initiatives.

Instead, big data and business analytics expertise should fall within existing functions—for example, finance, human resources, and marketing—with the aim of furthering the strategic initiatives of those functions. The efforts of the big data teams in these areas could be overseen and coordinated by a big data manager, reporting to the Chief Information Officer, who would

© 2014 World Economic Forum

### Box 2: Organizations already using big data initiatives

A few organizations that have followed frameworks for using big data include:

- A US-based mid- to upscale chain of department stores is gaining new insights from analyzing and combining data on Hadoop with data from traditional databases to turn its marketing staff from "Mad Men" to "Math Men."
- A US-based provider of business outsourcing solutions has set up an innovation lab where subject experts from different industries and backgrounds work together to tackle big data analytics.
- An Indianapolis-based global pharmaceutical company is using big data to develop an integrated approach to optimizing how clinical trials are conducted and eliminate inefficiencies.
- A US-based document management corporation is applying its decades of expertise in imaging technologies to transportation systems that can benefit from real-time analysis of data.
- Australia-based telecommunications companies use big data to determine which of their customers are less likely to pay their bills, allowing them to focus collection efforts on that group rather than across the whole customer base.
- A global corporation offering computer-assisted legal research services uses a big data technology platform it has developed in house both for its risk management business and for gathering data it sells to its clients. It now also sells this big data platform through its newly established subsidiary.
- A US-based multinational consumer goods company has developed a decision-support environment used by more than 60,000 employees worldwide to see what is happening in the business, to understand why it is happening, and to determine how to respond to changing market conditions. The decision cockpit is focused on forward-looking projections rather than historical reporting, with data visualization showing the relative revenue and profit contribution of each region, country, territory, brand, and product. The company's performance is also tracked against that of competing brands and products. The company has established about 50 collaborative conference rooms, called Business Spheres, in offices around the globe. The rooms are surrounded by projection screens for displaying the dashboards as well as live video-conferencing sessions, allowing remote executives to attend weekly review meetings in person.
- A California-based multinational energy corporation is using big data to transform the audit function. It runs audit tests on all of its accounts payable transactions instead of only on the small sample it used to analyze before. This enables the finance department to better understand various business risks and adjust audit coverage to the areas that pose the greatest risks. At the same time, it reduces the time spent on auditing by about 15 percent.
- A British multinational music recording and publishing company has created the Million Interview Dataset over the last few years, asking consumers in 24 countries and across 15 languages about their music listening and consumption habits. The Dataset provides rich insights into the interests, attitudes, behaviors, familiarity, and appreciation of music as expressed by music fans.
- A Minnesota-based member-owned agricultural cooperative combines 20 years of satellite imagery with local seed and crop protection data from its test sites to provide a service that generates field performance information for every acre and matches crop inputs and decisions to the potential of each field and each zone.

**Source:** HCL Technologies.

ensure that best practices were adopted and that initiatives were coordinated.

Following the nine steps described above will help the IT function to assume such responsibilities.

## CASE STUDIES

Many global organizations have already begun embarking on deriving value out of big data initiatives. Almost all of them have defined step-by-step frameworks somewhat similar to the one outlined above. The sheer variety of value creation evident—from clinical trials and marketing to risk management and audits, from analyzing crop and seed production to fan listening posts—is also staggering (see Box 2).

## RECOMMENDATIONS FOR GOVERNMENT ACTIONS AND POLICIES

In order to take full advantage of the potential of big data in both the public and private sectors, we recommend that governments create a vision and platform for public-sector open data. We believe that open data will be an *essential characteristic of future public policy.* It is important that such a vision percolate down from the top to garner support from ministries and civil servants alike so that open data initiatives function effectively.

Communicating from the very top that open data is an essential characteristic of public policy is crucial. Furthermore, governments should create an easy-to-use platform for the public to access the data in a form that is easily digestible and ready for analysis. It is also advisable to develop rules and regulations for taxing the commercial use of open data.

Governments should spearhead the effort to ensure the privacy and security of personal data. The appropriate agency should take a leading role in working with all relevant private- and public-sector entities to develop and implement policies for safeguarding personal data and means for enforcement.


© 2014 World Economic Forum

Moreover, it is essential to develop and execute a big data plan for all government services and activities. The plan should identify all government data worth analyzing, define data collection responsibilities, outline steps to ensure data quality, and determine where big data technologies and analysis capabilities should be first deployed.

Finally, each government should establish a big data center of excellence (BDCOE). The BDCOE should be the focal point of expertise, long-range thinking and policy formulation, and training and development. It should also be the repository of best practices. It should not only serve as a resource for all government agencies but should also act as the government's leading authority on all matters related to data management.

## CONCLUSION

Big data analytics is not a passing fad. It will be a central means of creating value for the organization of tomorrow—and that is "tomorrow" almost literally. It represents a major change in the way that businesses and other organizations will operate and will require a new mind-set and new capabilities. Given that, many organizations are struggling to know where to start in becoming competent in the realm of big data. A step-by-step approach can make the transition seem less daunting and minimize the stumbles that are bound to occur along the way.

## NOTES

1 Manyika et al. 2011.

2 Manyika et al. 2011.

3 Laney 2012.

4 Laney 2012.

5 HCL Technologies 2013a.

6 Manyika et al. 2011.

7 Manyika et al. 2011.

## REFERENCES

Gartner. 2013. *What's the Big Deal with Big Data for Customer Service?* Webinar with Gareth Herschel, Research Director, Gartner and Michael Maoz, VP Distinguished Analyst, Gartner. September 17. Available at http://my.gartner.com/portal/server.pt?open=512&objID=202&mode=2&PageID=5553&ref=webinar-rss&resId=2569818.

HCL Technologies. 2013a. Strategic Intelligence Wing Research on Big Data (the research arm of HCL Technologies).

——. 2013b. *CIO Straight Talk* Issue 3. Quincy, Mass, US and Noida, India: HCL Technologies. Available at http://magazine.straighttalkonline.com/issue3/.

Laney, D. 2012. *Big Data Strategy Components: IT Essentials.* October 15, ID G00238944. Chicago: Gartner.

Manyika, J., M. Chui, B. Brown, J. Bughin, R, Dobbs, C. Roxburgh, and A. Hung Byers. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." *McKinsey Global Institute Report.* May. Available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

© 2014 World Economic Forum