

HOW TO GET STARTED WITH BIG DATA

by Robert Souza, Rob Trollinger, Cornelius Kaestner, David Potere, and Jan Jamrich

THE USE OF “BIG DATA”—vast amounts of varied, fast-moving information—has the potential to magnify and accelerate the ability of businesses to understand customers and fine-tune products. Despite the availability of new techniques to make sense of big data, many senior leaders we encounter today are having a hard time figuring out where to start.

We see three overarching ways in which business leaders can unlock the value of big data:

- Develop a big-data strategy that capitalizes on a company’s most important data assets.
- Deploy more innovative advanced-analytical approaches to address the highest-priority challenges and processes involving big data.
- Determine how big-data transformation can improve existing business models and create entirely new revenue streams.

By far the quickest path to value over the short term is the second approach. By pushing the envelope in the fast-moving area of advanced analytics, companies will quickly learn what works best for them, where the value lies, and how to expand their capabilities over time. Such rapid learning can greatly inform the overall big-data strategy.

This article offers four questions to explore as you experiment with advanced-analytical approaches to big data. Answers to each one can help create immediate clarity about this seemingly vast topic.

Why Get Started Now?

For a long time, many executives thought it was difficult or dangerous to get into the big-data space. Burned by past experiments with large-scale customer-data initiatives that racked up excessive IT costs and failed to generate sufficient results, they have grown gun-shy about the topic. Bad memories of fizzled efforts have led some companies to get stuck in a state of “big-data paralysis.”

But these companies are overthinking things. It’s actually never been safer or easier to get started with big-data solutions. In fact, the rapid evolution of the field has made advanced analytics accessible to just about any company. Everything that a company needs in order to analyze large, complex data sets is now within arm’s reach. Three improvements stand out:

- *Increasing Speed and Decreasing Costs of Infrastructure and Hardware.* The amount of data that can pass through a fiber-optic cable continues to double every nine months. In parallel, processor and storage

costs have decreased by a factor of more than 1,000 over the past decade. (See Exhibit 1.)

- **Better Approaches to Utilizing Hardware.** Emerging technologies, in-memory processing, and accessible analytics platforms are enabling the processing of large data sets on thousands of servers distributed across the cloud. In addition, applications that can process big data sets now cost only a few thousand dollars rather than the millions they cost a decade ago. Now, customers can buy only what they need, when they need it.
- **Improved Availability of Technical Talent.** Finally, data analytic skills are becoming more widespread in the workforce, with major universities producing increasing numbers of graduate-degree holders who are focused on all aspects of advanced analytics.

How to Define Big Data

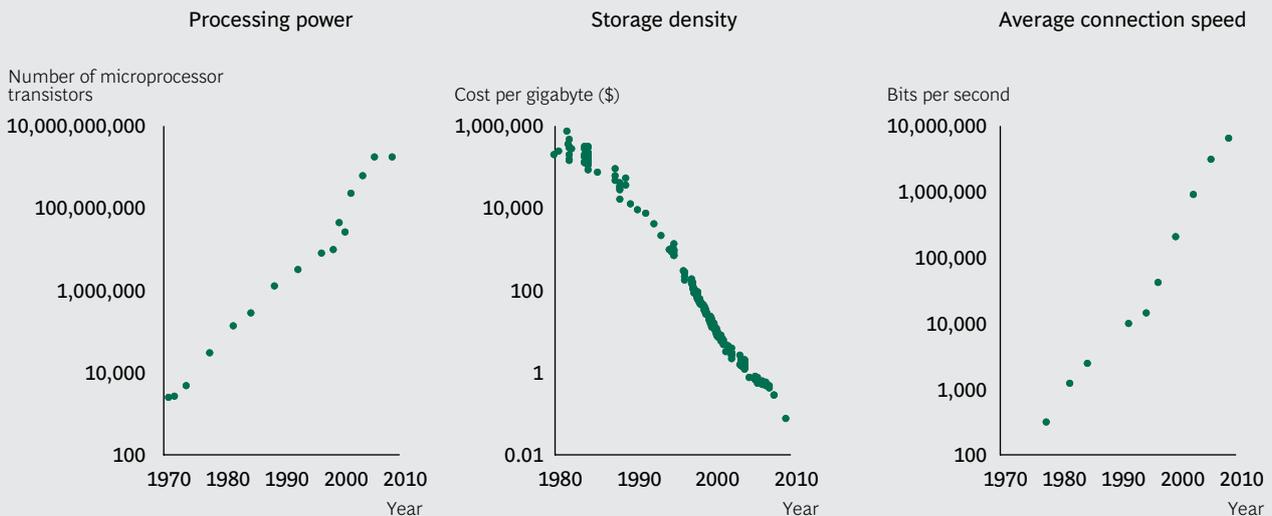
By now, leaders have no doubt heard the term “big data” used repeatedly in the media in different ways. Before we explore how to use advanced analytical techniques well, we must be clear about the meaning of big data.

Big data describes the large amounts and varieties of fast-moving information that can be processed and analyzed to create significant value. We’ll address three of the key characteristics first:

- **Volume.** Data that have grown to an immense size, prohibiting analysis with traditional tools
- **Variety.** Multiple formats of structured and unstructured data—such as social-media posts, location data from mobile devices, call center recordings, and sensor updates—that require fresh approaches to collection, storage, and management
- **Velocity.** Data that need to be processed in real or near-real time in order to be of greatest value, such as instantly providing a coupon to customers standing in the cereal aisle based on their past cereal purchases

Exhibit 2 shows the rapidly expanding nature of each of these three types of data. The three dimensions combine to create data sets that are often quite different from the traditional data a business collects about offers, purchases, and segments. The retail industry, for example, misses out on an estimated \$165 billion in total sales each year because retailers do not have the right products in stock to

EXHIBIT 1 | Recent Advances Have Made Big Data Accessible to a Much Broader Audience



Sources: Intel; ADB; IBM; archive.org; <http://ns1758.ca/winch/winchest.html>, courtesy of Matthew Komorowski; <http://www.mkomo.com/cost-per-gigabyte>.
Note: Internet speed is based on the average advertised download speed purchased by consumers, according to the CSG network (before 1994), the FCC (1994–2009), and the OECD (2011).

EXHIBIT 2 | The Volume, Variety, and Velocity of Data Are Expanding Exponentially

Data volume	Data volume increases every day at a rate of 2.5 exabytes, or 2.5 billion gigabytes <ul style="list-style-type: none">• The volume of Wal-Mart's hourly transaction data is 67 times that of what is contained in the Library of Congress• Smart meters in China alone will generate 13.3 trillion readings in 2020• Twelve terabytes of tweets are created each day
Data variety	Content comes from more-diverse sources and formats <ul style="list-style-type: none">• Data can come from smartphones, sensors, RFIDs, GPS devices, transactions, and more• Data can be in the form of databases, video, speech, social-media posts, links, tweets, text messages, and more
Data velocity	Faster analytics unlock value and reduce risk <ul style="list-style-type: none">• Organizations are analyzing millions of trading events per day to identify securities fraud• Organizations are analyzing hundreds of video feeds in real time to identify security threats

Source: BCG analysis.

meet customer demand. Big-data analysis allows companies to more quickly understand sales trends and incorporate more accurate forecasting, ultimately increasing customer loyalty and revenue.

Where Does the Opportunity Lie?

Many companies tend to struggle with a fourth dimension: value. Amid efforts to capture larger and more-diverse data sets, incorporate real-time social-media and location data, and keep up with the evolving technology landscape, leaders often lose perspective on how they can create value from advanced analytical approaches to big data.

Opportunities for value creation vary across industries. In retail, advanced analytical approaches often match well with strategies involving promotional effectiveness, pricing, store locations, and marketing at the individual level. In the energy industry, on the other hand, the emphasis is more often on making use of smart-meter data and optimizing physical assets, such as equipment and plants. In financial services, effective areas often include risk scoring, dynamic pricing, and finding optimal ATM and branch locations, while in insurance, the areas might include claims fraud, reimbursement optimization, and the tracking of driving behavior.

To identify value for your company, look for areas that have the following characteristics.

The volume of data matters. The outcome will sometimes be different if you analyze all the data as compared with sampling just a part. To target individual customers, a retailer needs to understand the entire purchase history of a given customer and how, for example, it is different from that of other customers. Using just a sample of customers or their transactions will lead to an incomplete picture, making promotional efforts less effective.

Consider the case of Cardlytics, an Atlanta-based startup that helps retailers sell to “markets of one.” Four of the top ten U.S. banks use the service to analyze hundreds of millions of customer transactions each week—a tremendous amount of data—in order to offer retailers the ability to customize promotions right on a customer’s bank statement. Because the offers are based on purchase behavior and where an individual customer shops, they are highly targeted. Merchants that use the service can reach customers, or competitors’ customers, in a simple, targeted way that can be tracked through actual usage. Response rates average 15 to 20 percent, as compared with the low single digits for most traditional campaigns. Consumers save money without having to print out coupons or enter promotion codes—the discounts are automatically credited to their statements. And banks earn extra revenue and offer their customers more rewards, without a heavy IT investment and without the data ever leaving the banks’ servers.

The variety of data matters. In some cases, the outcome will be different if a company analyzes diverse data types, ranging from structured data that can fit in a traditional relational database to the unstructured data that come from social-media posts and elsewhere and are difficult to map. When vast quantities of data are combined with fast-moving, unstructured social-media data, for instance, it can become very difficult to analyze with old-fashioned techniques.

As an example, there's tremendous value in accurately predicting churn at a customer-by-customer level at telecom companies. If a company offers discounts to people who would have stayed anyway, it has wasted its money. A lack of appropriate targeting can also make it overlook people who might leave for a competitor.

Companies can start with advanced analytics through teams, tools, and testing.

One company tackling this challenge is Telekomunikacja Polska, part of France Telecom-Orange Group and the largest fixed-line provider for voice and broadband services in Poland. The company wanted to quickly find ways to predict and address churn among its customers more effectively than the traditional methods, including analysis of declining rates of use and calculations of lifetime customer value based on how long customers stayed with the service and how much they spent.

The company decided to build a "social graph" from the call data records of millions of phone calls transiting through its network every month—looking, in particular, at patterns of who calls whom and at what frequency. The tool divides communities into roles such as "networkers," "bridges," "leaders," and "followers." For example, it detects the networkers, who link people together, and the leaders, who have a much greater impact on the network of people around them. That set of relational data gives the telecom-service provider much richer insight into who matters

among those who might drop its service and, therefore, how hard to try to keep its most valuable customers. As a result of the approach, the accuracy of the company's churn-prediction model has improved 47 percent.

The velocity of data matters. In other cases, companies need the latest real-time data to feed into decision making. Lack of that knowledge can mean an increase in risk. The faster a company reacts, the more likely it is to make a sale—or to prevent a customer from defecting to a competitor.

For example, insurers such as Allstate are beginning to offer pay-as-you-drive plans that incorporate a deluge of real-time customer-driving behavior gathered through devices installed in cars. Sensors measure how fast an individual customer drives, how much he or she drives, and how safely he or she navigates the roads. Signals of someone who is a higher risk can include hard braking, accelerating, and turning, as well as the number of miles driven at night or in rush-hour traffic.

Participating customers receive a discount on their next bill once they establish a baseline for their driving behavior over 30 days. They can also go online to track their performance, which tends to increase their loyalty. Likewise, insurers can avoid putting effort into retaining riskier customers, and they can increase these customer's rates to reflect what they learn about their overall risk pool.

Which Initial Steps to Take

Companies can get started with advanced analytics through what we call the "three Ts" of big-data effectiveness: teams, tools, and testing. Each offers an opportunity to start small, deliver tangible results, and scale up what works.

Build the right team. You need a SWAT team for analytics made up of well-rounded experts in the field. A diverse group of experts on narrow topics won't produce results quickly enough. A large company should start with a team of five, for a total people cost of roughly \$1.5 million. With each member of the team, companies need a combination of high-level analytical capabilities, technical familiarity with advanced-

analytics platforms, and a clear business perspective to discern which solutions are deployable and which are not. Team members don't have to be world-class on each dimension, but they do need strengths in each of the three skill sets. In many cases, the right combination of skills and experience is not available inside organizations. Partnerships often offer a shortcut to delivering value in such situations.

Deploy the right tools. You should next support these teams with the right tools to enable success. Each member should be able to leverage cloud infrastructure for his or her work. You can outfit each person with a virtual machine and massive amounts of storage for about \$15,000 per year. Many industry-standard tools cost only \$5,000 to \$15,000 per seat. The R open-source programming environment is free.

Test and learn the most effective approaches. Finally, run two- to three-month experiments that push for rapid results and implementation. This forces you to put a timeline on your analysis. You can't wait around for the perfect infrastructure or solution—the space is moving so quickly that the longer you wait, the further behind you'll be. If you're doing things right, you'll be learning what success means, what you discovered about your capabilities, and what kinds of infrastructure you need. In the process, you'll discover how resistant the organization is to taking advantage of advanced-analytical approaches to big data, which types of problems and data work best, and what other challenges you could apply your approaches to.

The experience of Aviva, the U.K.'s largest insurer, illustrates all three best practices. Creating Aviva's new Driving Behavior Data offering involved both developing algorithms to connect driving behavior with prices and creating a smartphone app. Instead of buying expensive systems, Aviva made a minimal initial investment and used a small number of developers. This enabled the team to get the first beta app to customers within five months. Over the next six months, the team refined the app and the customer experience based on rich data from thousands of driver journeys. In the process, it delivered a num-

ber of upgrades to the app—for example, enabling customers to share their scores on Twitter and Facebook. The additional knowledge about individual customers allows more-sophisticated insurance ratings based on how people actually drive rather than how similar types of people drive. Aviva can now offer better discounts to good drivers and create a more appealing product. So far, the app has been downloaded thousands of times, and plans are under way to deploy it in markets outside the U.K.

Or consider what has taken place at one North American food and beverage retailer. Not only does the company sell 30,000 individual items, but prices vary by location and market condition. And costs can change as often as four times per year. As a result, the retailer makes up to 120,000 price changes annually. The company, therefore, saw an opportunity to centralize its highly complex, high-data-volume pricing decisions. Using inexpensive tools and a team of only 11, the retailer was able to increase pricing accuracy and responsiveness and to deliver tens of millions of dollars in incremental sales and profit.

With these kinds of small steps, results are definitely possible. Smart companies that stick a paddle into the river of fast-moving data can begin to chart a direct course to creating significant value.

Robert Souza is a partner and managing director in the Boston office of The Boston Consulting Group. You may contact him by e-mail at souza.robert@bcg.com.

Rob Trollinger is a partner and managing director in the firm's Dallas office. You may contact him by e-mail at trollinger.rob@bcg.com.

Cornelius Kaestner is a principal in BCG's Washington office. You may contact him by e-mail at kaestner.cornelius@bcg.com.

David Potere is a principal in the firm's Boston office. You may contact him by e-mail at potere.david@bcg.com.

Jan Jamrich is a project leader in BCG's Boston office. You may contact him by e-mail at jamrich.jan@bcg.com.