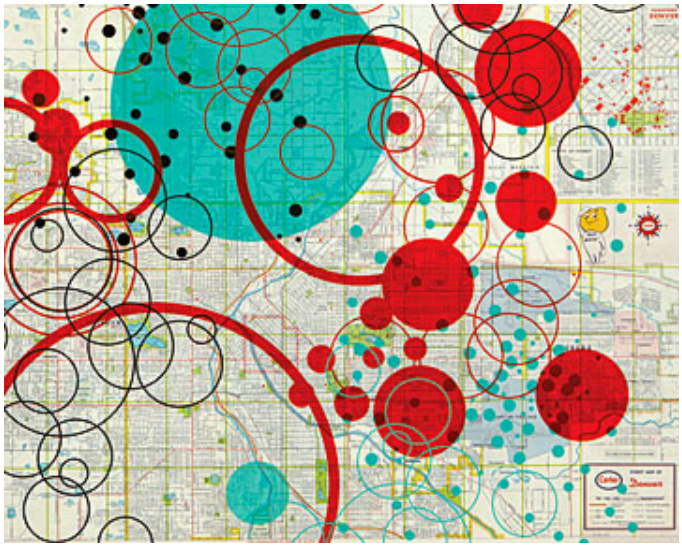
Harvard Business Review

DECISION MAKING

Big Data: The Management Revolution

by Andrew McAfee and Erik Brynjolfsson

FROM THE OCTOBER 2012 ISSUE



ARTWORK: TAMAR COHEN, HAPPY MOTORING, 2010, SILK SCREEN ON VINTAGE ROAD MAP, 26″ X 18″

"You can't manage what you don't measure."

There's much wisdom in that saying, which has been attributed to both W. Edwards Deming and Peter Drucker, and it explains why the recent explosion of digital data is so important. Simply put, because of big data, managers can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision making and performance.

Consider retailing. Booksellers in physical stores could always track which books sold and which did not. If they had a loyalty program, they could tie some of those purchases to individual customers. And that was about it. Once shopping moved online, though, the understanding of customers increased dramatically. Online retailers could track not only what customers bought, but also what else they looked at; how they navigated through the site; how much they were influenced by promotions, reviews, and page layouts; and similarities across individuals and groups. Before long, they developed algorithms to predict what books individual customers would like to read next—algorithms that performed better

every time the customer responded to or ignored a recommendation. Traditional retailers simply couldn't access this kind of information, let alone act on it in a timely manner. It's no wonder that Amazon has put so many brick-and-mortar bookstores out of business.

The familiarity of the Amazon story almost masks its power. We expect companies that were born digital to accomplish things that business executives could only dream of a generation ago. But in fact the use of big data has the potential to transform traditional businesses as well. It may offer them even greater opportunities for competitive advantage (online businesses have always known that they were competing on how well they understood their data). As we'll discuss in more detail, the big data of this revolution is far more powerful than the analytics that were used in the past. We can measure and therefore manage more precisely than ever before. We can make better predictions and smarter decisions. We can target more-effective interventions, and can do so in areas that so far have been dominated by gut and intuition rather than by data and rigor.

As the tools and philosophies of big data spread, they will change long-standing ideas about the value of experience, the nature of expertise, and the practice of management. Smart leaders across industries will see using big data for what it is: a management revolution. But as with any other major change in business, the challenges of becoming a big data-enabled organization can be enormous and require hands-on—or in some cases hands-off—leadership. Nevertheless, it's a transition that executives need to engage with today.

What's New Here?

Business executives sometimes ask us, "Isn't 'big data' just another way of saying 'analytics'?" It's true that they're related: The big data movement, like analytics before it, seeks to glean intelligence from data and translate that into business advantage. However, there are three key differences:

Volume.

As of 2012, about 2.5 exabytes of data are created each day, and that number is doubling every 40 months or so. More data cross the internet every second than were stored in the entire internet just 20 years ago. This gives companies an opportunity to work with many petabyes of data in a single data set—and not just from the internet. For instance, it is estimated that Walmart collects more than 2.5 petabytes of data every hour from its customer transactions. A petabyte is one quadrillion bytes, or the equivalent of about 20 million filing cabinets' worth of text. An exabyte is 1,000 times that amount, or one billion gigabytes.

Velocity.

For many applications, the speed of data creation is even more important than the volume. Real-time or nearly real-time information makes it possible for a company to be much more agile than its competitors. For instance, our colleague Alex "Sandy" Pentland and his group at the MIT Media Lab used location data from mobile phones to infer how many people were in Macy's parking lots on Black Friday—the start of the Christmas shopping season in the United States. This made it possible to estimate the retailer's sales on that critical day even before Macy's itself had recorded those sales. Rapid insights like that can provide an obvious competitive advantage to Wall Street analysts and Main Street managers.

Variety.

Big data takes the form of messages, updates, and images posted to social networks; readings from sensors; GPS signals from cell phones, and more. Many of the most important sources of big data are relatively new. The huge amounts of information from social networks, for example, are only as old as the networks themselves; Facebook was launched in 2004, Twitter in 2006. The same holds for smartphones and the other mobile devices that now provide enormous streams of data tied to people, activities, and locations. Because these devices are ubiquitous, it's easy to forget that the iPhone was unveiled only five years ago, and the iPad in 2010. Thus the structured databases that stored most corporate information until recently are ill suited to storing and processing big data. At the same time, the steadily declining costs of all the elements of computing—storage, memory, processing, bandwidth, and so on—mean that previously expensive data-intensive approaches are quickly becoming economical.

As more and more business activity is digitized, new sources of information and ever-cheaper equipment combine to bring us into a new era: one in which large amounts of digital information exist on virtually any topic of interest to a business. Mobile phones, online shopping, social networks, electronic communication, GPS, and instrumented machinery all produce torrents of data as a by-product of their ordinary operations. Each of us is now a walking data generator. The data available are often unstructured—not organized in a database—and unwieldy, but there's a huge amount of signal in the noise, simply waiting to be released. Analytics brought rigorous techniques to decision making; big data is at once simpler and more powerful. As Google's director of research, Peter Norvig, puts it: "We don't have better algorithms. We just have more data."

How Data-Driven Companies Perform

The second question skeptics might pose is this: "Where's the evidence that using big data intelligently will improve business performance?" The business press is rife with anecdotes and case studies that supposedly demonstrate the value of being data-driven. But the truth, we realized recently, is that nobody was tackling that question rigorously. To address this embarrassing gap, we led a team at the MIT Center for Digital Business, working in partnership with McKinsey's business technology office and with our colleague Lorin Hitt at Wharton and the MIT doctoral student Heekyung Kim. We set out to test the hypothesis that data-driven companies would be better performers. We conducted structured interviews with executives at 330 public North American companies about their organizational and technology management practices, and gathered performance data from their annual reports and independent sources.

Not everyone was embracing data-driven decision making. In fact, we found a broad spectrum of attitudes and approaches in every industry. But across all the analyses we conducted, one relationship stood out: The more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results. In particular, companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors. This performance difference remained robust after accounting for the

contributions of labor, capital, purchased services, and traditional IT investment. It was statistically significant and economically important and was reflected in measurable increases in stock market valuations.

Expertise from Surprising Sources

Often someone coming from outside an industry can spot a better way to use big data than an insider, just because so many new, unexpected sources of data are available. One of us, Erik, demonstrated this in research he conducted with Lynn Wu, now an assistant professor at Wharton. They used publicly available web search data to predict housing-price changes in metropolitan areas across the United States. They had no special knowledge of the housing market when they began their study, but they reasoned that virtually real-time search data would enable good near-term forecasts about the housing market—and they were right. In fact, their prediction proved more accurate than the official one from the National Association of Realtors, which had developed a far more complex model but relied on relatively slow-changing historical data.

This is hardly the only case in which simple models and big data trump more-elaborate analytics approaches. Researchers at the Johns Hopkins School of Medicine, for example, found that they could use data from Google Flu Trends (a free, publicly available aggregator of relevant search terms) to predict surges in flu-related emergency room visits a week before warnings came from the

So how are managers using big data? Let's look in detail at two companies that are far from Silicon Valley upstarts. One uses big data to create new businesses, the other to drive more sales.

Improved Airline ETAs

Minutes matter in airports. So does accurate information about flight arrival times: If a plane lands before the ground staff is ready for it, the passengers and crew are effectively trapped, and if it shows up later than expected, the staff sits idle, driving up costs. So when a major U.S. airline learned from an internal study that about 10% of the flights into its major hub had at least a 10-minute gap between the estimated time of arrival and the actual arrival time—and 30% had a gap of at least five minutes—it decided to take action.

At the time, the airline was relying on the aviation industry's long-standing practice of using the ETAs provided by pilots. The pilots made these estimates during their final approach to the airport, when they had many other demands on their time and attention. In search of a better solution, the airline turned to PASSUR Aerospace, a provider of decision-support technologies for the aviation industry. In 2001 PASSUR began offering its own arrival

estimates as a service called RightETA. It calculated these times by combining publicly available data about weather, flight schedules, and other factors with proprietary data the company itself collected, including feeds from a network of passive radar stations it had installed near airports to gather data about every plane in the local sky.

PASSUR started with just a few of these installations, but by 2012 it had more than 155. Every 4.6 seconds it collects a wide range of information about every plane that it "sees." This yields a huge and constant flood of digital data. What's more, the company keeps all the data it has gathered over time, so it has an immense body of multidimensional information spanning more than a decade. This allows sophisticated analysis and pattern matching. RightETA essentially works by asking itself "What happened all the previous times a plane approached this airport under these conditions? When did it actually land?"

After switching to RightETA, the airline virtually eliminated gaps between estimated and actual arrival times. PASSUR believes that enabling an airline to know when its planes are going to land and plan accordingly is worth several million dollars a year at each airport. It's a simple formula: Using big data leads to better predictions, and better predictions yield better decisions.

Speedier, More Personalized Promotions

A couple of years ago, Sears Holdings came to the conclusion that it needed to generate greater value from the huge amounts of customer, product, and promotion data it collected from its Sears, Craftsman, and Lands' End brands. Obviously, it would be valuable to combine and make use of all these data to tailor promotions and other offerings to customers, and to personalize the offers to take advantage of local conditions. Valuable, but difficult: Sears required about eight weeks to generate personalized promotions, at which point many of them were no longer optimal for the company. It took so long mainly because the data required for these large-scale analyses were both voluminous and highly fragmented—housed in many databases and "data warehouses" maintained by the various brands.

In search of a faster, cheaper way to do its analytic work, Sears Holdings turned to the technologies and practices of big data. As one of its first steps, it set up a Hadoop cluster. This is simply a group of inexpensive commodity servers whose activities are coordinated by an emerging software framework called Hadoop (named after a toy elephant in the household of Doug Cutting, one of its developers).

Sears started using the cluster to store incoming data from all its brands and to hold data from existing data warehouses. It then conducted analyses on the cluster directly, avoiding the time-consuming complexities of pulling data from various sources and combining them so that they can be analyzed. This change allowed the company to be much faster and more precise with its promotions. According to the company's CTO, Phil Shelley, the time needed to generate a comprehensive set of promotions dropped from eight weeks to one, and is still dropping. And these promotions are of higher quality, because they're more timely, more granular, and more personalized. Sears's Hadoop cluster stores and processes several petabytes of data at a fraction of the cost of a comparable standard data warehouse.

Shelley says he's surprised at how easy it has been to transition from old to new approaches to data management and high-performance analytics. Because skills and knowledge related to new data technologies were so rare in 2010, when Sears started the transition, it contracted some of the work to a company called Cloudera. But over time its old guard of IT and analytics professionals have become comfortable with the new tools and approaches.

The PASSUR and Sears Holding examples illustrate the power of big data, which allows more-accurate predictions, better decisions, and precise interventions, and can enable these things at seemingly limitless scale. We've seen big data used in supply chain management to understand why a carmaker's defect rates in the field suddenly increased, in customer service to continually scan and intervene in the health care practices of millions of people, in planning and forecasting to better anticipate online sales on the basis of a data set of product characteristics, and so on. We've seen similar payoffs in many other industries and functions, from finance to marketing to hotels and gaming, and from human resource management to machine repair.

Our statistical analysis tells us that what we're seeing is not just a few flashy examples but a more fundamental transformation of the economy. We've become convinced that almost no sphere of business activity will remain untouched by this movement.

A New Culture of Decision Making

The technical challenges of using big data are very real. But the managerial challenges are even greater—starting with the role of the senior executive team.

Muting the HiPPOs.

One of the most critical aspects of big data is its impact on how decisions are made and who gets to make them. When data are scarce, expensive to obtain, or not available in digital form, it makes sense to let well-placed people make decisions, which they do on the basis of experience they've built up and patterns and relationships they've observed and internalized. "Intuition" is the label given to this style of inference and decision making. People state their opinions about what the future holds—what's going to happen, how well something will work, and so on—and then plan accordingly. (See "The True Measures of Success," by Michael J. Mauboussin, in this issue.)

> Big data's power does not erase the need for vision or human insight.

For particularly important decisions, these people are typically high up in the organization, or they're expensive outsiders brought in because of their expertise and track records. Many in the big data community maintain that companies often make most of their important decisions by relying on "HiPPO"—the highest-paid person's opinion.

To be sure, a number of senior executives are genuinely data-driven and willing to override their own intuition when the data don't agree with it. But we believe that throughout the business world today, people rely too much on experience and intuition and not enough on

data. For our research we constructed a 5-point composite scale that captured the overall extent to which a company was data-driven. Fully 32% of our respondents rated their companies at or below 3 on this scale.

New roles.

Executives interested in leading a big data transition can start with two simple techniques. First, they can get in the habit of asking "What do the data say?" when faced with an important decision and following up with more-specific questions such as "Where did the data come from?," "What kinds of analyses were conducted?," and "How confident are we in the results?" (People will get the message quickly if executives develop this discipline.) Second, they can allow themselves to be overruled by the data; few things are more powerful for changing a decision-making culture than seeing a senior executive concede when data have disproved a hunch.

When it comes to knowing which problems to tackle, of course, domain expertise remains critical. Traditional domain experts—those deeply familiar with an area—are the ones who know where the biggest opportunities and challenges lie. PASSUR, for one, is trying to hire as many people as possible who have extensive knowledge of operations at America's major airports. They will be invaluable in helping the company figure out what offerings and markets it should go after next.

As the big data movement advances, the role of domain experts will shift. They'll be valued not for their HiPPO-style answers but because they know what questions to ask. Pablo Picasso might have been thinking of domain experts when he said, "Computers are useless. They can only give you answers."

Getting Started

You don't need to make enormous up-front investments in IT to use big data (unlike earlier generations of IT-enabled change). Here's one approach to building a capability from the ground up.

Five Management Challenges

Companies won't reap the full benefits of a transition to using big data unless they're able to manage change effectively. Five areas are particularly important in that process.

1. Pick a business unit to be the testing ground. It should have a quant-friendly leader backed up by a team of data scientists.

2. Challenge each key function to identify five business opportunities based on big data, each of which could be prototyped within five weeks by a team of no more than five people.

3. Implement a process for innovation that includes four steps: experimentation, measurement, sharing, and replication.

4. Keep in mind Joy's Law: "Most of the smartest people work for someone else." Open up some of your data sets and analytic challenges to interested parties across the internet and around the world.

Leadership.

Companies succeed in the big data era not simply because they have more or better data, but because they have leadership teams that set clear goals, define what success looks like, and ask the right questions. Big data's power does not erase the need for vision or human insight. On the contrary, we still must have business leaders who can spot a great opportunity, understand how a market is developing, think creatively and propose truly novel offerings, articulate a compelling vision, persuade people to embrace it and work hard to realize it, and deal effectively with customers, employees, stockholders, and other stakeholders. The successful companies of the next decade will be the ones whose leaders can do all that while changing the way their organizations make many decisions.

Talent management.

As data become cheaper, the complements to data become more valuable. Some of the most crucial of these are data scientists and other professionals skilled at working with large quantities of information. Statistics are important, but many of the key techniques for using big data are rarely taught in traditional statistics courses. Perhaps even more important are skills in cleaning and organizing large data sets; the new kinds of data rarely come in structured formats. Visualization tools and techniques are also increasing in value. Along with the data scientists, a new generation of computer scientists are bringing to bear techniques for working with very large data sets. Expertise in the design of experiments can help cross the gap between correlation and causation. The best data scientists are also comfortable speaking the language of business and helping leaders reformulate their

challenges in ways that big data can tackle. Not surprisingly, people with these skills are hard to find and in great demand. (See “Data Scientist: The Sexiest Job of the 21st Century,” by Thomas H. Davenport and D.J. Patil, in this issue.)

Technology.

The tools available to handle the volume, velocity, and variety of big data have improved greatly in recent years. In general, these technologies are not prohibitively expensive, and much of the software is open source. Hadoop, the most commonly used framework, combines commodity hardware with open-source software. It takes incoming streams of data and distributes them onto cheap disks; it also provides tools for analyzing the data. However, these technologies do require a skill set that is new to most IT departments, which will need to work hard to integrate all the relevant internal and external sources of data. Although attention to technology isn’t sufficient, it is always a necessary component of a big data strategy.

Decision making.

An effective organization puts information and the relevant decision rights in the same location. In the big data era, information is created and transferred, and expertise is often not where it used to be. The artful leader will create an organization flexible enough to minimize the “not invented here” syndrome and maximize cross-functional cooperation. People who understand the problems need to be brought together with the right data, but also with the people who have problem-solving techniques that can effectively exploit them.

Company culture.

The first question a data-driven organization asks itself is not “What do we think?” but “What do we know?” This requires a move away from acting solely on hunches and instinct. It also requires breaking a bad habit we’ve noticed in many organizations: pretending to be more data-driven than they actually are. Too often, we saw executives who spiced up their reports with lots of data that supported decisions they had already made using the traditional HiPPO approach. Only afterward were underlings dispatched to find the numbers that would justify the decision.Without question, many barriers to success remain. There are too few data scientists to go around. The technologies are new and in some cases exotic. It’s too easy to

mistake correlation for causation and to find misleading patterns in the data. The cultural challenges are enormous, and, of course, privacy concerns are only going to become more significant. But the underlying trends, both in the technology and in the business payoff, are unmistakable.

The evidence is clear: Data-driven decisions tend to be better decisions. Leaders will either embrace this fact or be replaced by others who do. In sector after sector, companies that figure out how to combine domain expertise with data science will pull away from their rivals. We can't say that all the winners will be harnessing big data to transform decision making. But the data tell us that's the surest bet.

A version of this article appeared in the October 2012 issue of *Harvard Business Review*.

Andrew McAfee is the co-director of the Initiative on the Digital Economy in the MIT Sloan School of Management. He is the author of *Enterprise 2.0* and the co-author, with Erik Brynjolfsson, of *The Second Machine Age*.

Erik Brynjolfsson is the Schussel Family Professor at MIT's Sloan School of Management and the director of its Center for Digital Business. They are the coauthors of *Race Against the Machine* (Digital Frontier Press, 2012).

This article is about DECISION MAKING

FOLLOW THIS TOPIC

Related Topics: COMPETITIVE STRATEGY | IT | INTERNET | DATA

Comments

Leave a Comment

POST

0 COMMENTS

JOIN THE CONVERSATION

POSTING GUIDELINES

We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.