

BUSINESS PROBLEM-SOLVING CASE

Does Big Data Bring Big Rewards?

Today's companies are dealing with an avalanche of data from social media, search, and sensors as well as from traditional sources. In 2012, an estimated 988 exabytes of digital information was generated, equivalent to a stack of books from the sun to Pluto and back. Making sense of "big data" has become one of the primary challenges for corporations of all shapes and sizes, but it also represents new opportunities. How are companies currently taking advantage of "big data?"

U.S. state and federal law enforcement agencies are analyzing big data to discover hidden patterns in criminal activity such as correlations between time, opportunity, and organizations, or non-obvious relationships (see Chapter 4) between individuals and criminal organizations that would be difficult to uncover in smaller data sets. New tools allow agencies to analyze data from a wide array of sources, including the Internet, and apply analytics to predict future crime patterns.

In New York City, the Real Time Crime Center data warehouse contains millions of data points on city crime and criminals. IBM and the New York Police Department (NYPD) worked together to create the warehouse, which contains data on over 120 million criminal complaints, 31 million national crime records, and 33 billion public records. The system's search capabilities allow the NYPD to quickly obtain data from any of these data sources. Information on criminals, such as a suspect's photo with details of past offenses or addresses with maps, can be visualized in seconds on a video wall or instantly relayed to officers at a crime scene.

Other organizations are using the data to go green, or, in the case of Vestas, to go even greener. Headquartered in Denmark, Vestas is the world's largest wind energy company, with over 43,000 wind turbines across 66 countries. Location data are important to Vestas so that it can accurately place its turbines for optimal wind power generation. Areas without enough wind will not generate the necessary power, but areas with too much wind may damage the turbines. Vestas relies on location-based data to determine the best spots to install their turbines.

To gather data on prospective turbine locations, Vestas's wind library combines data from global weather systems along with data from existing turbines. The company's previous wind library provided information in a grid pattern, with each grid measuring 27 x 27 kilometers (17x17 miles). Vestas engineers were able to bring the resolution down to about 10x10 meters (32x32 feet) to establish the exact wind flow pattern at a

particular location. To further increase the accuracy of its turbine placement models, Vestas needed to shrink the grid area even more, and this required ten times as much data as the previous system and a more powerful data management platform.

The company implemented a solution consisting of IBM InfoSphereBigInsights software running on a high-performance IBM System x iDataPlex server. (InfoSphereBigInsights is a set of software tools for "big data" analysis and visualization and is powered by Apache Hadoop.) Using these technologies, Vestas increased the size of its wind library and is able to manage and analyze location and weather data with models that are much more powerful and precise.

Vestas's wind library currently stores 2.8 petabytes of data and includes approximately 178 parameters, such as barometric pressure, humidity, wind direction, temperature, wind velocity, and other company historical data. Vestas plans to add global deforestation metrics, satellite images, geospatial data, and data on phases of the moon and tides.

The company can now reduce the resolution of its wind data grids by nearly 90 percent, down to a 3x3 kilometer area (about 1.8x1.8 miles). This capability enables Vestas to forecast optimal turbine placement in 15 minutes instead of three weeks, saving a month of development time for a turbine site and enabling Vestas customers to achieve a return on investment much more quickly.

AutoZone uses big data to help it adjust inventory and product prices at some of its 5,000 stores. For example, a customer walking into an AutoZone store in Waco, Texas, might find a deal on Gabriel shocks which that person would not find in most other AutoZone stores. The Mulberry, Florida AutoZone store might feature a special on a bug deflector. To target these deals at the local level, the auto parts retailer analyzes information gleaned from a variety of databases, such as the types of cars driven by people living around its retail outlets. Software from NuoDB, which uses a cloud services model, makes it possible to quickly increase the amount of data analyzed without bringing down the system, or changing a line of code.

Companies are also using big data solutions to analyze consumer sentiment. For example, car-rental giant Hertz gathers data from Web surveys, e-mails, text messages, Web site traffic patterns, and data generated at all of Hertz's 8,300 locations in 146 countries. The company now stores all of that data centrally instead of within each branch, reducing time

spent processing data and improving company response time to customer feedback and changes in sentiment. For example, by analyzing data generated from multiple sources, Hertz was able to determine that delays were occurring for returns in Philadelphia during specific times of the day. After investigating this anomaly, the company was able to quickly adjust staffing levels at its Philadelphia office during those peak times, ensuring a manager was present to resolve any issues. This enhanced Hertz's performance, and increased customer satisfaction.

There are limits to using big data. A number of companies have rushed to start big data projects without first establishing a business goal for this new information. Swimming in numbers doesn't necessarily mean that the right information is being collected or that people will make smarter decisions.

Several years ago, Google developed what it thought was a leading-edge algorithm using data it collected from Web searches to determine exactly how many people had influenza. It tried to calculate the number of people with flu in the United States by relating people's location to flu-related search queries on Google. According to Google Flu Trends, nearly 11 percent of the U.S. population was supposed to have had influenza at the flu season's peak in mid-January 2013. However, an article in the science journal *Nature* stated that Google's results were twice the actual amount estimated by the U.S. Centers for Disease Control and Prevention, which had 6 percent of the population coming down with the disease. Why did this happen? Several scientists suggested that Google was "tricked" by widespread media coverage of this year's severe flu season in the U.S., which was further amplified by social media coverage. Google's algorithm only looked at numbers, not the context of the search results.

Sears Holdings, the parent company of Sears and Kmart, is trying to use big data to get closer to its customers. Sears used to be the largest retailer in the United States, but for many years has steadily lost ground to discounters such as Walmart and Target and to competitively-priced specialty retailers such as Home Depot and Lowe's. The company has been slow to reduce operating costs, keep pace with current merchandising trends, and remodel its 2,173 U.S. stores, many of which are run-down and in undesirable locations.

Over the years, Sears had invested heavily in information technology. At one time it spent more on information technology and networking than all other non-computer firms in the United States except the Boeing Corporation. Sears used its huge customer databases of 60 million past and present Sears credit card holders to target groups such as tool buyers, appliance buyers, and gardening enthusiasts with special promotions. These efforts did not translate into competitive

advantage because Sears's cost structure remained one of the highest in its industry.

The Sears company has continued to embrace new technology to revive flagging sales: online shopping, mobile apps, and an Amazon.com-like marketplace with other vendors for 18 million products, along with heavy in-store promotions. So far, these efforts have not paid off, and sales have declined since the 2005 merger with Kmart. The company posted a loss of \$930 million for 2012.

Sears Holdings CEO Lou D'Ambrosio thinks that even more intensive use of technology and mining of customer data will be the answer. The expectation is that deeper knowledge of customer preferences and buying patterns will make promotions, merchandising, and selling much more effective. Customers will flock to Sears stores because they will be carrying exactly what they want.

A customer loyalty program called Shop Your Way Rewards promises customers generous free deals for repeat purchases if they agree to share their personal shopping data with the company. Sears would not disclose how many customers have signed up for Shop Your Way Rewards, but loyalty-marketing firm Colloquy estimates around 50 million people are members.

Sears wants to personalize marketing campaigns, coupons, and offers down to the individual customer, but its legacy systems were incapable of supporting that level of activity. In order to use big models on large data sets, Sears turned to Apache Hadoop and big data technology. It used to take Sears six weeks to analyze marketing campaigns for loyalty club members using a mainframe, Teradata data warehouse software, and SAS servers. Using Hadoop, the processing can be completed weekly. Certain online and mobile commerce analyses can be performed daily and targeting is much more precise, in some cases down to the individual customer. Sears's old models were able to use 10 percent of available data, but the new models are able to work with 100 percent. In the past, Sears was only able to retain data from 90 days to two years, but with Hadoop, it can keep everything, increasing its chances of finding more meaningful patterns in the data.

What's more, Hadoop processing is much less costly than conventional relational databases. A Hadoop system handling 200 terabytes of data runs about one-third the cost of a 200-terabyte relational platform. With Hadoop's massively parallel processing power, processing 2 billion records takes Sears little more than one minute longer than processing 100 million records.

Hadoop is still an immature platform, and Hadoop expertise is scarce. Sears had to learn Hadoop largely by trial and error. But it now runs critical reports on the platform, including analyses of customers, financial data, products, and supply chains. Capitalizing on its

experience as a big-data innovator, Sears set up a subsidiary called MetaScale to sell big data cloud and consulting services to other companies.

Sears can point to many conceptual uses of Hadoop, but the question still lingers about whether the company is effectively using Hadoop to solve its enormous business problems. Is it truly able to offer customers personalized promotions and are they working? What is the business impact? Where are the numbers to show that big data is helping Sears become more profitable? Sears may be able to generate revenue by selling big data expertise to MetaScale customers, but will Hadoop really help turn Sears around?

Jim Sullivan, a partner at loyalty marketing firm Colloquy, notes that a good loyalty program that gives a company better intelligence about what its customers really want can be a strategic advantage, but even the best loyalty programs can't fix a fundamentally broken brand.

Sources: Rachael King and Steven Rosenbush, "Big Data Broadens Its Range," *The Wall Street Journal* (March 13, 2013); Nick Bilton, "Disruptions: Data Without a Context Tells a Misleading Story," *The New York Times*, February 24, 2013; Shira Ovide, "Big Data, Big Blunders," *The Wall Street Journal*, March 11, 2013; Mark A. Smith, "Big Data Pointless without Integration," *Information Management*,

February 25, 2013; Frank Konkel, "Fast Failure Could Lead to Big-Data Success," *Federal Computer Week*, January 30, 2013; Doug Henschen, "Why Sears Is Going All-in on Hadoop," *Information Week*, October 3, 2012; Samuel Greengard, "Big Data Unlocks Business Value," *Baseline*, January 2012; Paul S. Barth, "Managing Big Data: What Every CIO Needs to Know," *CIO Insight*, January 12, 2012; IBM Corporation, "Vestas: Turning Climate into Capital with Big Data," 2011; IBM Corporation, "Extending and Enhancing Law Enforcement Capabilities" and "How Big Data Is Giving Hertz a Big Advantage," 2010.

Case Study Questions

- 6-14** Describe the kinds of "big data" collected by the organizations described in this case.
- 6-15** List and describe the business intelligence technologies described in this case.
- 6-16** Why did the companies described in this case need to maintain and analyze big data? What business benefits did they obtain? How much were they helped by analyzing big data?
- 6-17** Identify three decisions that were improved by using big data.
- 6-18** Should all organizations try to analyze big data? Why or why not? What people, organization, and technology issues should be addressed before a company decides to work with big data?