



This is “The Data Asset: Databases, Business Intelligence, and Competitive Advantage”, chapter 11 from the book [Getting the Most Out of Information Systems \(index.html\)](#) (v. 1.4).

This book is licensed under a [Creative Commons by-nc-sa 3.0](http://creativecommons.org/licenses/by-nc-sa/3.0/) license. See the license for more details, but that basically means you can share this book as long as you credit the author (but see below), don't make money from it, and do make it available to everyone else under the same terms.

This content was accessible as of December 29, 2012, and it was downloaded then by [Andy Schmitz](#) (<http://lardbucket.org>) in an effort to preserve the availability of this book.

Normally, the author and publisher would be credited here. However, the publisher has asked for the customary Creative Commons attribution to the original publisher, authors, title, and book URI to be removed. Additionally, per the publisher's request, their name has been removed in some passages. More information is available on this project's [attribution page](http://2012books.lardbucket.org/attribution.html?utm_source=header).

For more information on the source of this book, or why it is available for free, please see [the project's home page](#) (<http://2012books.lardbucket.org/>). You can browse or download additional books there.

Chapter 11

The Data Asset: Databases, Business Intelligence, and Competitive Advantage

11.1 Introduction

LEARNING OBJECTIVES

1. Understand how increasingly standardized data, access to third-party data sets, cheap, fast computing and easier-to-use software are collectively enabling a new age of decision making.
2. Be familiar with some of the enterprises that have benefited from data-driven, fact-based decision making.

The planet is awash in data. Cash registers ring up transactions worldwide. Web browsers leave a trail of cookie crumbs nearly everywhere they go. And with radio frequency identification (RFID), inventory can literally announce its presence so that firms can precisely journal every hop their products make along the value chain: “I’m arriving in the warehouse,” “I’m on the store shelf,” “I’m leaving out the front door.”

A study by Gartner Research claims that the amount of data on corporate hard drives doubles every six months. C. Babcock, “Data, Data, Everywhere”, *InformationWeek*, January 9, 2006. while IDC states that the collective number of those bits already exceeds the number of stars in the universe. L. Mearian, “Digital Universe and Its Impact Bigger Than We Thought,” *Computerworld*, March 18, 2008. Wal-Mart alone boasts a data volume well over 125 times as large as the entire print collection of the U.S. Library of Congress, and rising. Derived by comparing Wal-Mart’s 2.5 petabytes (E. Lai, “Teradata Creates Elite Club for Petabyte-Plus Data Warehouse Customers,” *Computerworld*, October 18, 2008) to the Library of Congress estimate of 20 TB (D. Gewirtz, “What If Someone Stole the Library of Congress?” *CNN.com/AC360*, May 25, 2009). It’s further noted that the Wal-Mart figure is just for data stored on systems provided by the vendor Teradata. Wal-Mart has many systems outside its Teradata-sourced warehouses, too. You’ll hear managers today broadly refer to this torrent of bits as “**Big Data**”¹.

And with this flood of data comes a tidal wave of opportunity. Increasingly standardized corporate data, and access to rich, third-party data sets—all leveraged by cheap, fast computing and easier-to-use software—are collectively enabling a new age of data-driven, fact-based decision making. You’re less likely to hear old-school terms like “decision support systems” used to describe what’s going on here. The phrase of the day is **business intelligence (BI)**², a catchall term combining aspects of reporting, data exploration and ad hoc queries, and sophisticated data

1. A general term used to describe massive amount of data available to today’s managers. Big data are often unstructured and are too big and costly to easily work through use of conventional databases, but new tools are making these massive datasets available for analysis and insight.
2. A term combining aspects of reporting, data exploration and ad hoc queries, and sophisticated data modeling and analysis.

modeling and analysis. Alongside business intelligence in the new managerial lexicon is the phrase **analytics**³, a term describing the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. T. Davenport and J. Harris, *Competing on Analytics: The New Science of Winning* (Boston: Harvard Business School Press, 2007).

The benefits of all this data and number crunching are very real, indeed. Data leverage lies at the center of competitive advantage we've studied in the Zara, Netflix, and Google cases. Data mastery has helped vault Wal-Mart to the top of the *Fortune* 500 list. It helped Harrah's Casino Hotels grow to be twice as profitable as similarly sized Caesars and rich enough to acquire this rival (Harrah's did decide that it liked the Caesars name better and is now known as Caesars Entertainment). And data helped Capital One find valuable customers that competitors were ignoring, delivering ten-year financial performance a full ten times greater than the S&P 500. Data-driven decision making is even credited with helping the Red Sox win their first World Series in eighty-three years and with helping the New England Patriots win three Super Bowls in four years. To quote from a *BusinessWeek* cover story on analytics, "Math Will Rock Your World!" S. Baker, "Math Will Rock Your World," *BusinessWeek*, January 23, 2006, http://www.businessweek.com/magazine/content/06_04/b3968001.htm.

Sounds great, but it can be a tough slog getting an organization to the point where it has a leveragable data asset. In many organizations data lies dormant, spread across inconsistent formats and incompatible systems, unable to be turned into anything of value. Many firms have been shocked at the amount of work and complexity required to pull together an infrastructure that empowers its managers. But not only can this be done, it must be done. Firms that are basing decisions on hunches aren't managing; they're gambling. And today's markets have no tolerance for uninformed managerial dice rolling.

While we'll study technology in this chapter, our focus isn't as much on the technology itself as it is on what you can do with that technology. Consumer products giant P&G believes in this distinction so thoroughly that the firm renamed its IT function as "Information and Decision Solutions." J. Soat, "P&G's CIO Puts IT at Users' Service," *InformationWeek*, December 15, 2007. Solutions drive technology decisions, not the other way around.

3. A term describing the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.

In this chapter we'll study the data asset, how it's created, how it's stored, and how it's accessed and leveraged. We'll also study many of the firms mentioned above, and more; providing a context for understanding how managers are leveraging data

to create winning models, and how those that have failed to realize the power of data have been left in the dust.

Data, Analytics, and Competitive Advantage

Anyone can acquire technology—but data is oftentimes considered a defensible source of competitive advantage. The data a firm can leverage is a true strategic asset when it's rare, valuable, imperfectly imitable, and lacking in substitutes (see [Chapter 2 "Strategy and Technology: Concepts and Frameworks for Understanding What Separates Winners from Losers"](#)).

If more data brings more accurate modeling, moving early to capture this rare asset can be the difference between a dominating firm and an also-ran. But be forewarned, there's no monopoly on math. Advantages based on capabilities and data that others can acquire will be short-lived. Those advances leveraged by the Red Sox were originally pioneered by the Oakland A's and are now used by nearly every team in the major leagues.

This doesn't mean that firms can ignore the importance data can play in lowering costs, increasing customer service, and other ways that boost performance. But differentiation will be key in distinguishing operationally effective data use from those efforts that can yield true strategic positioning.

KEY TAKEAWAYS

- The amount of data on corporate hard drives doubles every six months.
- In many organizations, available data is not exploited to advantage.
- Data is oftentimes considered a defensible source of competitive advantage; however, advantages based on capabilities and data that others can acquire will be short-lived.

QUESTIONS AND EXERCISES

1. Name and define the terms that are supplanting discussions of decision support systems in the modern IS lexicon.
2. Is data a source of competitive advantage? Describe situations in which data might be a source for sustainable competitive advantage. When might data not yield sustainable advantage?
3. Are advantages based on analytics and modeling potentially sustainable? Why or why not?
4. What role do technology and timing play in realizing advantages from the data asset?

11.2 Data, Information, and Knowledge

LEARNING OBJECTIVES

1. Understand the difference between data and information.
2. Know the key terms and technologies associated with data organization and management.

Data⁴ refers simply to raw facts and figures. Alone it tells you nothing. The real goal is to turn data into **information**⁵. Data becomes information when it's presented in a context so that it can answer a question or support decision making. And it's when this information can be combined with a manager's **knowledge**⁶—their insight from experience and expertise—that stronger decisions can be made.

4. Raw facts and figures.
5. Data presented in a context so that it can answer a question or support decision making.
6. Insight derived from experience and expertise.

Trusting Your Data

The ability to look critically at data and assess its validity is a vital managerial skill. When decision makers are presented with wrong data, the results can be disastrous. And these problems can get amplified if bad data is fed to automated systems. As an example, look at the series of man-made and computer-triggered events that brought about a billion-dollar collapse in United Airlines stock.

In the wee hours one Sunday morning in September 2008, a single reader browsing back stories on the *Orlando Sentinel's* Web site viewed a 2002 article on the bankruptcy of United Airlines (UAL went bankrupt in 2002, but emerged from bankruptcy four years later). That lone Web surfer's access of this story during such a low-traffic time was enough for the *Sentinel's* Web server to briefly list the article as one of the paper's "most popular." Google crawled the site and picked up this "popular" news item, feeding it into Google News.

Early that morning, a worker in a Florida investment firm came across the Google-fed story, assumed United had yet again filed for bankruptcy, then posted a summary on Bloomberg. Investors scanning Bloomberg jumped on what looked like a reputable early warning of another United bankruptcy, dumping UAL stock. Blame the computers again—the rapid plunge from these early trades caused automatic sell systems to kick in (event-triggered, computer-automated trading is responsible for about 30 percent of all stock trades). Once the machines took over, UAL dropped like a rock, falling from twelve to three dollars. That drop represented the vanishing of \$1 billion in wealth, and all this because no one checked the date on a news story. Welcome to the new world of paying attention! M. Harvey, "Probe into How Google Mix-Up Caused \$1 Billion Run on United," *Times Online*, September 12, 2008, http://technology.timesonline.co.uk/tol/news/tech_and_web/article4742147.ece.

Understanding How Data Is Organized: Key Terms and Technologies

A **database**⁷ is simply a list (or more likely, several related lists) of data. Most organizations have several databases—perhaps even hundreds or thousands. And these various databases might be focused on any combination of functional areas (sales, product returns, inventory, payroll), geographical regions, or business units.

7. A single table or a collection of related tables.

Firms often create specialized databases for recording transactions, as well as databases that aggregate data from multiple sources in order to support reporting and analysis.

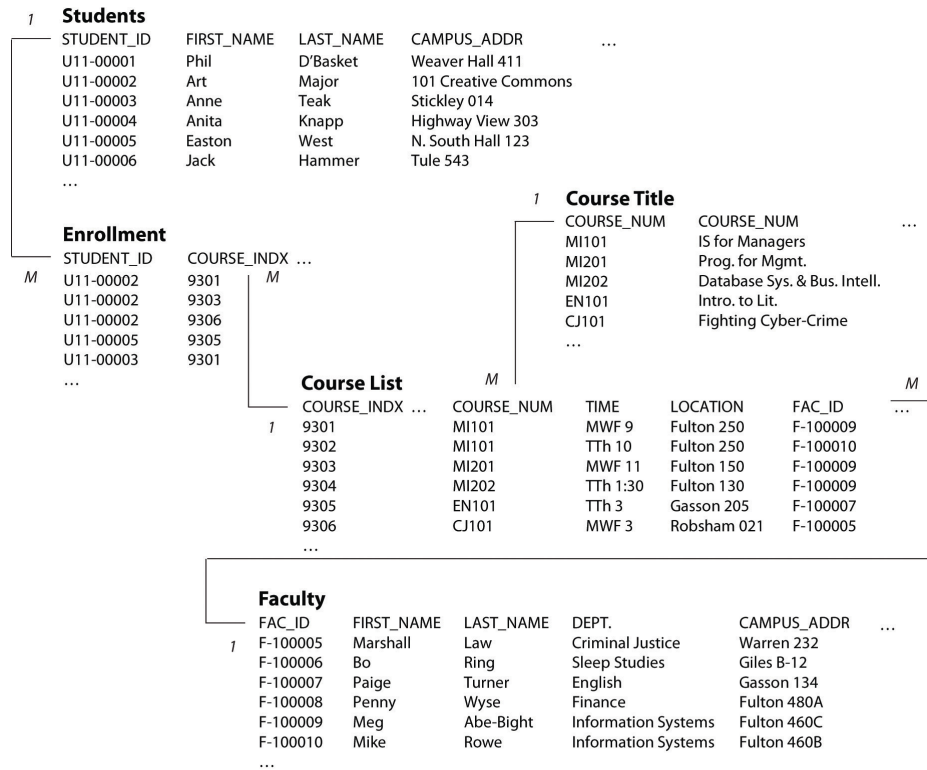
Databases are created, maintained, and manipulated using programs called **database management systems (DBMS)**⁸, sometimes referred to as *database software*. DBMS products vary widely in scale and capabilities. They include the single-user, desktop versions of Microsoft Access or Filemaker Pro, Web-based offerings like Intuit QuickBase, and industrial strength products from Oracle, IBM (DB2), Sybase, Microsoft (SQL Server), and others. Oracle is the world's largest database software vendor, and database software has meant big bucks for Oracle cofounder and CEO Larry Ellison. Ellison perennially ranks in the Top 10 of the *Forbes* 400 list of wealthiest Americans.

The acronym SQL (often pronounced *sequel*) also shows up a lot when talking about databases. **Structured query language (SQL)**⁹ is by far the most common language for creating and manipulating databases. You'll find variants of SQL inhabiting everything from lowly desktop software, to high-powered enterprise products. Microsoft's high-end database is even called SQL Server. And of course there's also the open source MySQL (whose stewardship now sits with Oracle as part of the firm's purchase of Sun Microsystems). Given this popularity, if you're going to learn one language for database use, SQL's a pretty good choice. And for a little inspiration, visit Monster.com or another job site and search for jobs mentioning SQL. You'll find page after page of listings, suggesting that while database systems have been good for Ellison, learning more about them might be pretty good for you, too.

8. Sometimes called "database software"; software for creating, maintaining, and manipulating data.
9. A language used to create and manipulate databases.
10. Job title focused on directing, performing, or overseeing activities associated with a database or set of databases. These may include (but not necessarily be limited to): database design, creation, implementation, maintenance, backup and recovery, policy setting and enforcement, and security.

Even if you don't become a database programmer or **database administrator (DBA)**¹⁰, you're almost surely going to be called upon to dive in and use a database. You may even be asked to help identify your firm's data requirements. It's quite common for nontech employees to work on development teams with technical staff, defining business problems, outlining processes, setting requirements, and determining the kinds of data the firm will need to leverage. Database systems are powerful stuff, and can't be avoided, so a bit of understanding will serve you well.

Figure 11.1 A Simplified Relational Database for a University Course Registration System



A complete discourse on technical concepts associated with database systems is beyond the scope of our managerial introduction, but here are some key concepts to help get you oriented, and that all managers should know.

- A **table or file**¹¹ refers to a list of data.
- A *database* is either a single table or a collection of related tables. The course registration database above depicts five tables.
- A **column or field**¹² defines the data that a table can hold. The “Students” table above shows columns for STUDENT_ID, FIRST_NAME, LAST_NAME, CAMPUS_ADDR (the “...” symbols above are meant to indicate that in practice there may be more columns or rows than are shown in this simplified diagram).
- A **row or record**¹³ represents a single instance of whatever the table keeps track of. In the example above, each row of the “Students” table represents a student, each row of the “Enrollment” table represents the enrollment of a student in a particular course, and each row of the “Course List” represents a given section of each course offered by the University.

11. A list of data, arranged in columns (fields) and rows (records).

12. A column in a database table. Columns represent each category of data contained in a record (e.g., first name, last name, ID number, date of birth).

13. A row in a database table. Records represent a single instance of whatever the table keeps track of (e.g., student, faculty, course title).

- A **key**¹⁴ is the field used to relate tables in a database. Look at how the STUDENT_ID key is used above. There is *one* unique STUDENT_ID for each student, but the STUDENT_ID may appear *many* times in the “Enrollment” table, indicating that each student may be enrolled in many classes. The “1” and “M” in the diagram above indicate the one to many relationships among the keys in these tables.

Databases organized like the one above, where multiple tables are related based on common keys, are referred to as **relational databases**¹⁵. There are many other database formats (sporting names like *hierarchical*, and *object-oriented*), but relational databases are far and away the most popular. And all SQL databases are relational databases.

Even though SQL and the relational model are hugely popular and dominate many corporate environments, other systems exist. An increasingly popular set of technologies known as NoSQL avoid SQL and the rigid structure of relational databases. NoSQL technologies are especially popular with Internet firms that rely on massive, unwieldy, and disparately structured data.

We’ve just scratched the surface for a very basic introduction. Expect that a formal class in database systems will offer you far more detail and better design principles than are conveyed in the elementary example above. But you’re already well on your way!

KEY TAKEAWAYS

- Data includes raw facts that must be turned into information in order to be useful and valuable.
- Databases are created, maintained, and manipulated using programs called database management systems (DBMS), sometimes referred to as database software.
- All data fields in the same database have unique names, several data fields make up a data record, multiple data records make up a table or data file, and one or more tables or data files make up a database.
- Relational databases are the most common database format.

14. Code that unlocks encryption.

15. The most common standard for expressing databases, whereby tables (files) are related based on common keys.

QUESTIONS AND EXERCISES

1. Define the following terms: table, record, field. Provide another name for each term along with your definition.
2. Answer the following questions using the course registration database system, diagramed above:
 - a. Imagine you also want to keep track of student majors. How would you do this? Would you modify an existing table? Would you add new tables? Why or why not?
 - b. Why do you suppose the system needs a “Course Title” table?
 - c. This database is simplified for our brief introduction. What additional data would you need to keep track of if this were a real course registration system? What changes would you make in the database above to account for these needs?
3. Research to find additional examples of organizations that made bad decisions based on bad data. Report your examples to your class. What was the end result of the examples you’re citing (e.g., loss, damage, or other outcome)? What could managers have done to prevent problems in the cases that you cited? What role did technology play in the examples that you cite? What role did people or procedural issues play?
4. Why is an understanding of database terms and technologies important, even for nontechnical managers and staff? Consider factors associated with both system use and system development. What other skills, beyond technology, may be important when engaged in data-driven decision making?

11.3 Where Does Data Come From?

LEARNING OBJECTIVES

1. Understand various internal and external sources for enterprise data.
2. Recognize the function and role of data aggregators, the potential for leveraging third-party data, the strategic implications of relying on externally purchased data, and key issues associated with aggregators and firms that leverage externally sourced data.

Organizations can pull together data from a variety of sources. While the examples that follow aren't meant to be an encyclopedic listing of possibilities, they will give you a sense of the diversity of options available for data gathering.

Transaction Processing Systems

For most organizations that sell directly to their customers, **transaction processing systems (TPS)**¹⁶ represent a fountain of potentially insightful data. Every time a consumer uses a point-of-sale system, an ATM, or a service desk, there's a **transaction**¹⁷ (some kind of business exchange) occurring, representing an event that's likely worth tracking.

The cash register is the data generation workhorse of most physical retailers, and the primary source that feeds data to the TPS. But while TPS can generate a lot of bits, it's sometimes tough to match this data with a specific customer. For example, if you pay a retailer in cash, you're likely to remain a mystery to your merchant because your name isn't attached to your money. Grocers and retailers can tie you to cash transactions if they can convince you to use a **loyalty card**¹⁸. Use one of these cards and you're in effect giving up information about yourself in exchange for some kind of financial incentive. The explosion in retailer cards is directly related to each firm's desire to learn more about you and to turn you into a more loyal and satisfied customer.

Some cards provide an instant discount (e.g., the CVS Pharmacy ExtraCare card), while others allow you to build up points over time (Best Buy's Reward Zone). The latter has the additional benefit of acting as a switching cost. A customer may think "I could get the same thing at Target, but at Best Buy, it'll increase my existing points balance and soon I'll get a cash back coupon."

16. Systems that record a transaction (some form of business-related exchange), such as a cash register sale, ATM withdrawal, or product return.
17. Some kind of business exchange.
18. Systems that provide rewards and usage incentives, typically in exchange for a method that provides a more detailed tracking and recording of customer activity. In addition to enhancing data collection, loyalty cards can represent a significant switching cost.

Tesco: Tracked Transactions, Increased Insights, and Surging Sales

UK grocery giant Tesco, the planet's third-largest retailer, is envied worldwide for what analysts say is the firm's unrivaled ability to collect vast amounts of retail data and translate this into sales. K. Capell, "Tesco: 'Wal-Mart's Worst Nightmare,'" *BusinessWeek*, December 29, 2008.

Tesco's data collection relies heavily on its ClubCard loyalty program, an effort pioneered back in 1995. But Tesco isn't just a physical retailer. As the world's largest Internet grocer, the firm gains additional data from Web site visits, too. Remove products from your virtual shopping cart? Tesco can track this. Visited a product comparison page? Tesco watches which product you've chosen to go with and which you've passed over. Done your research online, then traveled to a store to make a purchase? Tesco sees this, too.

Tesco then mines all this data to understand how consumers respond to factors such as product mix, pricing, marketing campaigns, store layout, and Web design. Consumer-level targeting allows the firm to tailor its marketing messages to specific subgroups, promoting the right offer through the right channel at the right time and the right price. To get a sense of Tesco's laser-focused targeting possibilities, consider that the firm sends out close to ten million different, targeted offers each quarter. T. Davenport and J. Harris, "Competing with Multichannel Marketing Analytics," *Advertising Age*, April 2, 2007. Offer redemption rates are the best in the industry, with some coupons scoring an astronomical 90 percent usage! M. Lowenstein, "Tesco: A Retail Customer Divisibility Champion," *CustomerThink*, October 20, 2002.

The firm's data-driven management is clearly delivering results. Even while operating in the teeth of a global recession, Tesco repeatedly posted record corporate profits and the highest earnings ever for a British retailer. K. Capell, "Tesco Hits Record Profit, but Lags in U.S.," *BusinessWeek*, April 21, 2009; A. Hawkes, "Tesco Reports Record Profits of £3.8bn," *Guardian*, April 19, 2011.

Enterprise Software (CRM, SCM, and ERP)

Firms increasingly set up systems to gather additional data beyond conventional purchase transactions or Web site monitoring. CRM or customer relationship management systems are often used to empower employees to track and record

data at nearly every point of customer contact. Someone calls for a quote? Brings a return back to a store? Writes a complaint e-mail? A well-designed CRM system can capture all these events for subsequent analysis or for triggering follow-up events.

Enterprise software includes not just CRM systems but also categories that touch every aspect of the value chain, including supply chain management (SCM) and enterprise resource planning (ERP) systems. More importantly, enterprise software tends to be more integrated and standardized than the prior era of proprietary systems that many firms developed themselves. This integration helps in combining data across business units and functions, and in getting that data into a form where it can be turned into information (for more on enterprise systems, see [Chapter 9 "Understanding Software: A Primer for Managers"](#)).

Surveys

Sometimes firms supplement operational data with additional input from surveys and focus groups. Oftentimes, direct surveys can tell you what your cash register can't. Zara store managers informally survey customers in order to help shape designs and product mix. Online grocer FreshDirect (see [Chapter 2 "Strategy and Technology: Concepts and Frameworks for Understanding What Separates Winners from Losers"](#)) surveys customers weekly and has used this feedback to drive initiatives from reducing packaging size to including star ratings on produce. R. Braddock, "Lessons of Internet Marketing from FreshDirect," *Wall Street Journal*, May 11, 2009. Many CRM products also have survey capabilities that allow for additional data gathering at all points of customer contact.

Can Technology “Cure” U.S. Health Care?

The U.S. health care system is broken. It’s costly, inefficient, and problems seem to be getting worse. Estimates suggest that health care spending makes up a whopping 18 percent of U.S. gross domestic product. J. Zhang, “Recession Likely to Boost Government Outlays on Health Care,” *Wall Street Journal*, February 24, 2009. U.S. automakers spend more on health care than they do on steel. S. Milligan, “Business Warms to Democratic Leaders,” *Boston Globe*, May 28, 2009. Even more disturbing, it’s believed that medical errors cause as many as ninety-eight thousand unnecessary deaths in the United States each year, more than motor vehicle accidents, breast cancer, or AIDS. R. Appleton, “Less Independent Doctors Could Mean More Medical Mistakes,” *InjuryBoard.com*, June 14, 2009; and B. Obama, President’s Speech to the American Medical Association, Chicago, IL, June 15, 2009, http://www.whitehouse.gov/the_press_office/Remarks-by-the-President-to-the-Annual-Conference-of-the-American-Medical-Association.

For years it’s been claimed that technology has the potential to reduce errors, improve health care quality, and save costs. Now pioneering hospital networks and technology companies are partnering to help tackle cost and quality issues. For a look at possibilities for leveraging data throughout the doctor-patient value chain, consider the “event-driven medicine” system built by Dr. John Halamka and his team at Boston’s Beth Israel Deaconess Medical Center (part of the Harvard Medical School network).

When docs using Halamka’s system encounter a patient with a chronic disease, they generate a decision support “screening sheet.” Each event in the system: an office visit, a lab results report (think the medical equivalent of transactions and customer interactions), updates the patient database. Combine that electronic medical record information with **artificial intelligence**¹⁹ on best practice, and the system can offer recommendations for care, such as, “Patient is past due for an eye exam” or, “Patient should receive pneumovax [a vaccine against infection] this season.” J. Halamka, “IT Spending: When Less Is More,” *BusinessWeek*, March 2, 2009. The systems don’t replace decision making by doctors and nurses, but they do help to ensure that key issues are on a provider’s radar.

More efficiencies and error checks show up when prescribing drugs. Docs are presented with a list of medications covered by that patient’s insurance,

19. Computer software that seeks to reproduce or mimic (perhaps with improvements) human thought, decision making, or brain functions.

allowing them to choose quality options while controlling costs. Safety issues, guidelines, and best practices are also displayed. When correct, safe medication in the right dose is selected, the electronic prescription is routed to the patients' pharmacy of choice. As Halamka puts it, going from "doctor's brain to patients vein" without any of that messy physician handwriting, all while squeezing out layers where errors from human interpretation or data entry might occur.

President Obama believes technology initiatives can save health care as much as \$120 billion a year, or roughly two thousand five hundred dollars per family. D. McCullagh, "Q&A: Electronic Health Records and You," *CNET/CBSNews.com*, May 19, 2009. An aggressive number, to be sure. But with such a large target to aim at, it's no wonder that nearly every major technology company now has a health solutions group. Microsoft and Google even offer competing systems for electronically storing and managing patient health records. If systems like Halamka's and others realize their promise, big benefits may be just around the corner.

External Sources

Sometimes it makes sense to combine a firm's data with bits brought in from the outside. Many firms, for example, don't sell directly to consumers (this includes most drug companies and packaged goods firms). If your firm has partners that sell products for you, then you'll likely rely heavily on data collected by others.

Data bought from sources available to all might not yield competitive advantage on its own, but it can provide key operational insight for increased efficiency and cost savings. And when combined with a firm's unique data assets, it may give firms a high-impact edge.

Consider restaurant chain Brinker, a firm that runs seventeen hundred eateries in twenty-seven countries under the Chili's, On The Border, and Maggiano's brands. Brinker (whose ticker symbol is EAT), supplements their own data with external feeds on weather, employment statistics, gas prices, and other factors, and uses this in predictive models that help the firm in everything from determining staffing levels to switching around menu items. R. King, "Intelligence Software for Business," *BusinessWeek* podcast, February 27, 2009.

In another example, Carnival Cruise Lines combines its own customer data with third-party information tracking household income and other key measures. This data plays a key role in a recession, since it helps the firm target limited marketing dollars on those past customers that are more likely to be able to afford to go on a cruise. So far it's been a winning approach. For three years in a row, the firm has experienced double-digit increases in bookings by repeat customers. R. King, "Intelligence Software for Business," *BusinessWeek* podcast, February 27, 2009.

Who's Collecting Data about You?

There's a thriving industry collecting data about you. Buy from a catalog, fill out a warranty card, or have a baby, and there's a very good chance that this event will be recorded in a database somewhere, added to a growing digital dossier that's made available for sale to others. If you've ever gotten catalogs, coupons, or special offers from firms you've never dealt with before, this was almost certainly a direct result of a behind-the-scenes trafficking in the "digital you."

Firms that trawl for data and package them up for resale are known as **data aggregators**²⁰. They include Acxiom, a \$1.3 billion a year business that combines public source data on real estate, criminal records, and census reports, with private information from credit card applications, warranty card surveys, and magazine subscriptions. The firm holds data profiling some two hundred million Americans. A. Gefter and T. Simonite, "What the Data Miners Are Digging Up about You," *CNET*, December 1, 2008.

Or maybe you've heard of Lexis-Nexis. Many large universities subscribe to the firm's electronic newspaper, journal, and magazine databases. But the firm's parent, Reed Elsevier, is a data sales giant, with divisions packaging criminal records, housing information, and additional data used to uncover corporate fraud and other risks. In February, 2008, the firm got even more data rich, acquiring Acxiom competitor ChoicePoint for \$4.1 billion. With that kind of money involved, it's clear that data aggregation is very big business. A. Greenberg, "Companies That Profit from Your Data," *Forbes*, May 14, 2008.

The Internet also allows for easy access to data that had been public but otherwise difficult to access. For one example, consider home sale prices and home value assessments. While technically in the public record, someone wanting this information previously had to traipse down to their Town Hall and speak to a clerk, who would hand over a printed log book. Not exactly a Google-speed query. Contrast this with a visit to Zillow.com. The free site lets you pull up a map of your town and instantly peek at how much your neighbors paid for their homes. And it lets them see how much you paid for yours, too.

Computerworld's Robert Mitchell uncovered a more disturbing issue when public record information is made available online. His New Hampshire municipality had digitized and made available some of his old public documents without

20. Firms that collect and resell data.

obscuring that holy grail for identity thieves, his Social Security number. R. Mithchell, “Why You Should Be Worried about Your Privacy on the Web,” *Computerworld*, May 11, 2009.

Then there are accuracy concerns. A record incorrectly identifying you as a cat lover is one thing, but being incorrectly named to the terrorist watch list is quite another. During a five-week period airline agents tried to block a particularly high profile U.S. citizen from boarding airplanes on five separate occasions because his name resembled an alias used by a suspected terrorist. That citizen? The late Ted Kennedy, who at the time was the senior U.S. senator from Massachusetts. R. Swarns, “Senator? Terrorist? A Watch List Stops Kennedy at Airport,” *New York Times*, August 20, 2004.

For the data trade to continue, firms will have to treat customer data as the sacred asset it is. Step over that “creep-out” line, and customers will push back, increasingly pressing for tighter privacy laws. Data aggregator Intellius used to track cell phone customers, but backed off in the face of customer outrage and threatened legislation.

Another concern—sometimes data aggregators are just plain sloppy, committing errors that can be costly for the firm and potentially devastating for victimized users. For example, in 2005, ChoicePoint accidentally sold records on 145,000 individuals to a cybercrime identity theft ring. The ChoicePoint case resulted in a \$15 million fine from the Federal Trade Commission. A. Greenberg, “Companies That Profit from Your Data,” *Forbes*, May 14, 2008. In 2011, hackers stole at least 60 million e-mail addresses from marketing firm Epsilon, prompting firms as diverse as Best Buy, Citi, Hilton, and the College Board to go through the time-consuming, costly, and potentially brand-damaging process of warning customers of the breach. Epsilon faces liabilities charges of almost a quarter of a billion dollars, but some estimate that the total price tag for the breach could top \$4 billion. F. Rashid, “Epsilon Data Breach to Cost Billions in Worst-Case Scenario,” *eWeek*, May 3, 2011. Just because you can gather data and traffic in bits doesn’t mean that you should. Any data-centric effort should involve input not only from business and technical staff, but from the firm’s legal team as well (for more, see the box “[Note 11.33 "Privacy Regulation: A Moving Target"](#)").

Privacy Regulation: A Moving Target

New methods for tracking and gathering user information appear daily, testing user comfort levels. For example, the firm Umbria uses software to analyze millions of blog and forum posts every day, using sentence structure, word choice, and quirks in punctuation to determine a blogger's gender, age, interests, and opinions. While Google refused to include facial recognition as an image search product ("too creepy," said its chairman), M. Warman, "Google Warns against Facial Recognition Database," *Telegraph*, May 16, 2011. Facebook, with great controversy, turned on facial recognition by default. N. Bilton, "Facebook Changes Privacy Settings to Enable Facial Recognition," *New York Times*, June 7, 2011. It's quite possible that in the future, someone will be able to upload a photo to a service and direct it to find all the accessible photos and video on the Internet that match that person's features. And while targeting is getting easier, a Carnegie Mellon study showed that it doesn't take much to find someone with a minimum of data. Simply by knowing gender, birth date, and postal zip code, 87 percent of people in the United States could be pinpointed by name. A. Gefter and T. Simonite, "What the Data Miners Are Digging Up about You," *CNET*, December 1, 2008. Another study showed that publicly available data on state and date of birth could be used to predict U.S. Social Security numbers—a potential gateway to identity theft. E. Mills, "Report: Social Security Numbers Can Be Predicted," *CNET*, July 6, 2009, http://news.cnet.com/8301-1009_3-10280614-83.html.

Some feel that Moore's Law, the falling cost of storage, and the increasing reach of the Internet have us on the cusp of a privacy train wreck. And that may inevitably lead to more legislation that restricts data-use possibilities. Noting this, strategists and technologists need to be fully aware of the legal environment their systems face (see [Chapter 14 "Google in Three Parts: Search, Online Advertising, and Beyond"](#) for examples and discussion) and consider how such environments may change in the future. Many industries have strict guidelines on what kind of information can be collected and shared.

For example, HIPAA (the U.S. Health Insurance Portability and Accountability Act) includes provisions governing data use and privacy among health care providers, insurers, and employers. The financial industry has strict requirements for recording and sharing communications between firm and client (among many other restrictions). There are laws limiting the kinds of information that can be gathered on younger Web surfers. And there are several laws operating at the state level as well.

International laws also differ from those in the United States. Europe, in particular, has a strict European Privacy Directive. The directive includes governing provisions that limit data collection, require notice and approval of many types of data collection, and require firms to make data available to customers with mechanisms for stopping collection efforts and correcting inaccuracies at customer request. Data-dependent efforts plotted for one region may not fully translate in another effort if the law limits key components of technology use. The constantly changing legal landscape also means that what works today might not be allowed in the future.

Firms beware—the public will almost certainly demand tighter controls if the industry is perceived as behaving recklessly or inappropriately with customer data.

KEY TAKEAWAYS

- For organizations that sell directly to their customers, transaction processing systems (TPS) represent a source of potentially useful data.
- Grocers and retailers can link you to cash transactions if they can convince you to use a loyalty card which, in turn, requires you to give up information about yourself in exchange for some kind of financial incentive such as points or discounts.
- Enterprise software (CRM, SCM, and ERP) is a source for customer, supply chain, and enterprise data.
- Survey data can be used to supplement a firm's operational data.
- Data obtained from outside sources, when combined with a firm's internal data assets, can give the firm a competitive edge.
- Data aggregators are part of a multibillion-dollar industry that provides genuinely helpful data to a wide variety of organizations.
- Data that can be purchased from aggregators may not in and of itself yield sustainable competitive advantage since others may have access to this data, too. However, when combined with a firm's proprietary data or integrated with a firm's proprietary procedures or other assets, third-party data can be a key tool for enhancing organizational performance.
- Data aggregators can also be quite controversial. Among other things, they represent a big target for identity thieves, are a method for spreading potentially incorrect data, and raise privacy concerns.
- Firms that mismanage their customer data assets risk lawsuits, brand damage, lower sales, fleeing customers, and can prompt more restrictive legislation.
- Further raising privacy issues and identity theft concerns, recent studies have shown that in many cases it is possible to pinpoint users through allegedly anonymous data, and to guess Social Security numbers from public data.
- New methods for tracking and gathering user information are raising privacy issues which possibly will be addressed through legislation that restricts data use.

QUESTIONS AND EXERCISES

1. Why would a firm use a loyalty card? What is the incentive for the firm? What is the incentive for consumers to opt in and use loyalty cards? What kinds of strategic assets can these systems create?
2. In what ways does Tesco gather data? Can other firms match this effort? What other assets does Tesco leverage that helps the firm remain among top performing retailers worldwide?
3. Make a list of the kind of data you might give up when using a cash register, a Web site, or a loyalty card, or when calling a firm's customer support line. How might firms leverage this data to better serve you and improve their performance?
4. Are you concerned by any of the data-use possibilities that you outlined in prior questions, discussed in this chapter, or that you've otherwise read about or encountered? If you are concerned, why? If not, why not? What might firms, governments, and consumers do to better protect consumers?
5. What are some of the sources data aggregators tap to collect information?
6. Privacy laws are in a near constant state of flux. Conduct research to identify the current state of privacy law. Has major legislation recently been proposed or approved? What are the implications for firms operating in effected industries? What are the potential benefits to consumers? Do consumers lose anything from this legislation?
7. Self-regulation is often proposed as an alternative to legislative efforts. What kinds of efforts would provide "teeth" to self-regulation. Are there steps firms could do to make you believe in their ability to self-regulate? Why or why not?
8. What is HIPPA? What industry does it impact?
9. How do international privacy laws differ from U.S. privacy laws?

11.4 Data Rich, Information Poor

LEARNING OBJECTIVES

1. Know and be able to list the reasons why many organizations have data that can't be converted to actionable information.
2. Understand why transactional databases can't always be queried and what needs to be done to facilitate effective data use for analytics and business intelligence.
3. Recognize key issues surrounding data and privacy legislation.

Despite being awash in data, many organizations are data rich but information poor. A survey by consulting firm Accenture found 57 percent of companies reporting that they didn't have a beneficial, consistently updated, companywide analytical capability. Among major decisions, only 60 percent were backed by analytics—40 percent were made by intuition and gut instinct. R. King, "Business Intelligence Software's Time Is Now," *BusinessWeek*, March 2, 2009. The big culprit limiting BI initiatives is getting data into a form where it can be used, analyzed, and turned into information. Here's a look at some factors holding back information advantages.

Incompatible Systems

Just because data is collected doesn't mean it can be used. This limit is a big problem for large firms that have **legacy systems**²¹, outdated information systems that were not designed to share data, aren't compatible with newer technologies, and aren't aligned with the firm's current business needs. The problem can be made worse by mergers and acquisitions, especially if a firm depends on operational systems that are incompatible with its partner. And the elimination of incompatible systems isn't just a technical issue. Firms might be under extended agreement with different vendors or outsourcers, and breaking a contract or invoking an escape clause may be costly. Folks working in M&A (the area of investment banking focused on valuing and facilitating mergers and acquisitions) beware—it's critical to uncover these hidden costs of technology integration before deciding if a deal makes financial sense.

21. Older information systems that are often incompatible with other systems, technologies, and ways of conducting business. Incompatible legacy systems can be a major roadblock to turning data into information, and they can inhibit firm agility, holding back operational and strategic initiatives.

Legacy Systems: A Prison for Strategic Assets

The experience of one *Fortune* 100 firm that your author has worked with illustrates how incompatible information systems can actually hold back strategy. This firm was the largest in its category, and sold identical commodity products sourced from its many plants worldwide. Being the biggest should have given the firm scale advantages. But many of the firm's manufacturing facilities and international locations developed or purchased separate, incompatible systems. Still more plants were acquired through acquisition, each coming with its own legacy systems.

The plants with different information systems used *different* part numbers and naming conventions even though they sold *identical* products. As a result, the firm had no timely information on how much of a particular item was sold to which worldwide customers. The company was essentially operating as a collection of smaller, regional businesses, rather than as the worldwide behemoth that it was.

After the firm developed an information system that standardized data across these plants, it was, for the first time, able to get a single view of worldwide sales. The firm then used this data to approach their biggest customers, negotiating lower prices in exchange for increased commitments in worldwide purchasing. This trade let the firm take share from regional rivals. It also gave the firm the ability to shift manufacturing capacity globally, as currency prices, labor conditions, disaster, and other factors impacted sourcing. The new information system in effect liberated the latent strategic asset of scale, increasing sales by well over a billion and a half dollars in the four years following implementation.

Operational Data Can't Always Be Queried

Another problem when turning data into information is that most transactional databases aren't set up to be simultaneously accessed for reporting and analysis. When a customer buys something from a cash register, that action may post a sales record and deduct an item from the firm's inventory. In most TPS systems, requests made to the database can usually be performed pretty quickly—the system adds or modifies the few records involved and it's done—in and out in a flash.

But if a manager asks a database to analyze historic sales trends showing the most and least profitable products over time, they may be asking a computer to look at thousands of transaction records, comparing results, and neatly ordering findings. That's not a quick in-and-out task, and it may very well require significant processing to come up with the request. Do this against the very databases you're using to record your transactions, and you might grind your computers to a halt.

Getting data into systems that can support analytics is where data warehouses and data marts come in, the topic of our next section.

KEY TAKEAWAYS

- A major factor limiting business intelligence initiatives is getting data into a form where it can be used (i.e., analyzed and turned into information).
- Legacy systems often limit data utilization because they were not designed to share data, aren't compatible with newer technologies, and aren't aligned with the firm's current business needs.
- Most transactional databases aren't set up to be simultaneously accessed for reporting and analysis. In order to run analytics the data must first be ported to a data warehouse or data mart.

QUESTIONS AND EXERCISES

1. How might information systems impact mergers and acquisitions? What are the key issues to consider?
2. Discuss the possible consequences of a company having multiple plants, each with a different information system using different part numbers and naming conventions for identical products.
3. Why does it take longer, and require more processing power, to analyze sales trends by region and product, as opposed to posting a sales transaction?

11.5 Data Warehouses and Data Marts

LEARNING OBJECTIVES

1. Understand what data warehouses and data marts are and the purpose they serve.
2. Know the issues that need to be addressed in order to design, develop, deploy, and maintain data warehouses and data marts.

Since running analytics against transactional data can bog down a system, and since most organizations need to combine and reformat data from multiple sources, firms typically need to create separate data repositories for their reporting and analytics work—a kind of staging area from which to turn that data into information.

Two terms you'll hear for these kinds of repositories are **data warehouse**²² and **data mart**²³. A data warehouse is a set of databases designed to support decision making in an organization. It is structured for fast online queries and exploration. Data warehouses may aggregate enormous amounts of data from many different operational systems.

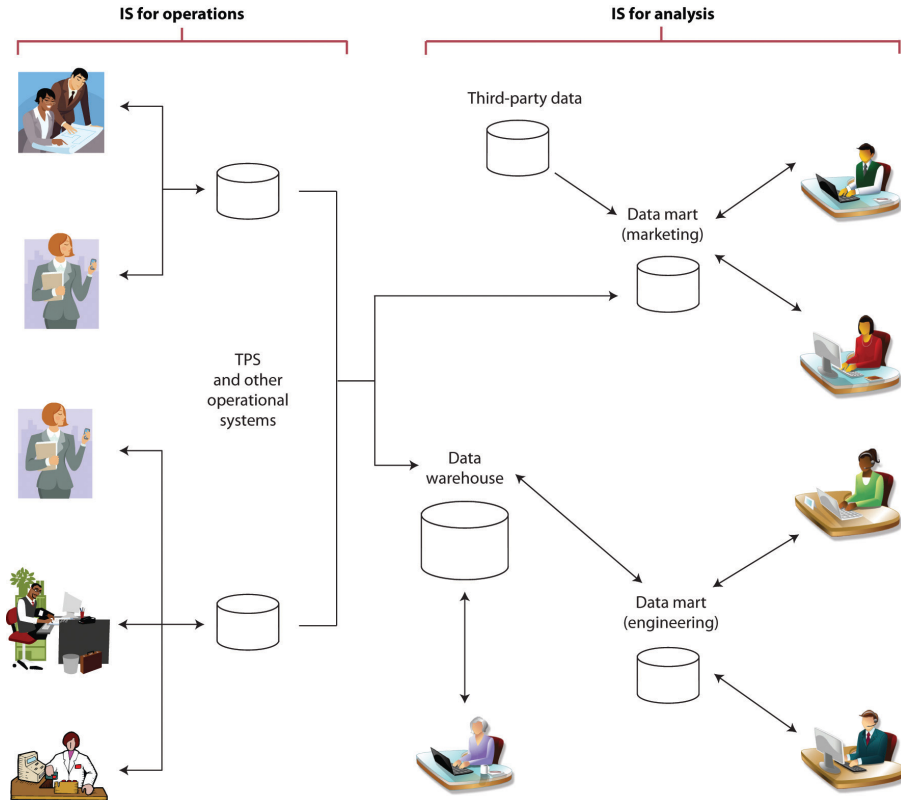
A data mart is a database focused on addressing the concerns of a specific problem (e.g., increasing customer retention, improving product quality) or business unit (e.g., marketing, engineering).

Marts and warehouses may contain huge volumes of data. For example, a firm may not need to keep large amounts of historical point-of-sale or transaction data in its operational systems, but it might want past data in its data mart so that managers can hunt for patterns and trends that occur over time.

22. A set of databases designed to support decision making in an organization.

23. A database or databases focused on addressing the concerns of a specific problem (e.g., increasing customer retention, improving product quality) or business unit (e.g., marketing, engineering).

Figure 11.2



Information systems supporting operations (such as TPS) are typically separate, and “feed” information systems used for analytics (such as data warehouses and data marts).

It’s easy for firms to get seduced by a software vendor’s demonstration showing data at your fingertips, presented in pretty graphs. But as mentioned earlier, getting data in a format that can be used for analytics is hard, complex, and challenging work. Large data warehouses can cost millions and take years to build. Every dollar spent on technology may lead to five to seven more dollars on consulting and other services. R. King, “Intelligence Software for Business,” *BusinessWeek* podcast, February 27, 2009.

Most firms will face a trade-off—do we attempt a large-scale integration of the whole firm, or more targeted efforts with quicker payoffs? Firms in fast-moving industries or with particularly complex businesses may struggle to get sweeping projects completed in enough time to reap benefits before business conditions change. Most consultants now advise smaller projects with narrow scope driven by specific business goals. D. Rigby and D. Ledingham, “CRM Done Right,” *Harvard*

Business Review, November 2004; and R. King, “Intelligence Software for Business,” *BusinessWeek* podcast, February 27, 2009.

Firms can eventually get to a unified data warehouse but it may take time. Even analytics king Wal-Mart is just getting to that point. Retail giant Wal-Mart once reported having over seven hundred different data marts and hired Hewlett-Packard for help in bringing the systems together to form a more integrated data warehouse. H. Havenstein, “HP Nabs Wal-Mart as Data Warehousing Customer,” *Computerworld*, August 1, 2007.

The old saying from the movie *Field of Dreams*, “If you build it, they will come,” doesn’t hold up well for large-scale data analytics projects. This work should start with a clear vision with business-focused objectives. When senior executives can see objectives illustrated in potential payoff, they’ll be able to champion the effort, and experts agree, having an executive champion is a key success factor. Focusing on business issues will also drive technology choice, with the firm better able to focus on products that best fit its needs.

Once a firm has business goals and hoped-for payoffs clearly defined, it can address the broader issues needed to design, develop, deploy, and maintain its system: Key points adapted from Davenport and J. Harris, *Competing on Analytics: The New Science of Winning* (Boston: Harvard Business School Press, 2007).

- *Data relevance*. What data is needed to compete on analytics and to meet our current and future goals?
- *Data sourcing*. Can we even get the data we’ll need? Where can this data be obtained from? Is it available via our internal systems? Via third-party data aggregators? Via suppliers or sales partners? Do we need to set up new systems, surveys, and other collection efforts to acquire the data we need?
- *Data quantity*. How much data is needed?
- *Data quality*. Can our data be trusted as accurate? Is it clean, complete, and reasonably free of errors? How can the data be made more accurate and valuable for analysis? Will we need to ‘scrub,’ calculate, and consolidate data so that it can be used?
- *Data hosting*. Where will the systems be housed? What are the hardware and networking requirements for the effort?
- *Data governance*. What rules and processes are needed to manage data from its creation through its retirement? Are there operational issues (backup, disaster recovery)? Legal issues? Privacy issues? How should the firm handle security and access?

For some perspective on how difficult this can be, consider that an executive from one of the largest U.S. banks once lamented at how difficult it was to get his systems to do something as simple as properly distinguishing between men and women. The company's customer-focused data warehouse drew data from thirty-six separate operational systems—bank teller systems, ATMs, student loan reporting systems, car loan systems, mortgage loan systems, and more. Collectively these legacy systems expressed gender in *seventeen* different ways: “M” or “F”; “m” or “f”; “Male” or “Female”; “MALE” or “FEMALE”; “1” for man, “0” for woman; “0” for man, “1” for woman and more, plus various codes for “unknown.” The best math in the world is of no help if the values used aren't any good. There's a saying in the industry, “garbage in, garbage out.”

Hadoop: Big Insights from Unstructured “Big Data”

Having neatly structured data warehouses and data-marts are great—the tools are reliable and can often be turned over to end-users or specialists who can rapidly produce reports and other analyses. But roughly 80 percent of corporate data is messy and unstructured, and it is not stored in conventional, relational formats—think of data stored in office productivity documents, e-mail, and social media. R. King, “Getting a Handle on Big Data with Hadoop,” *BusinessWeek*, September 7, 2011. Conventional tools often choke when trying to sift through the massive amounts of data collected by many of today's firms. The open-source project known as Hadoop was created to analyze massive amounts of raw information better than traditional, highly-structured databases.

Hadoop is made up of some half-dozen separate software pieces and requires the integration of these pieces to work. Hadoop-related projects have names such as Hive, Pig, and Zookeeper. Their use is catching on like wildfire, with some expecting that within five years, more than half of the world's data will be stored in Hadoop environments. R. King, “Getting a Handle on Big Data with Hadoop,” *BusinessWeek*, September 7, 2011. Expertise is in short supply, with Hadoop-savvy technologists having lots of career opportunities.

There are four primary advantages to Hadoop: IBM Big Data, “What is Hadoop?” YouTube video, 3:12 May 22, 2012, <http://www.youtube.com/watch?v=RQr0qd8gxW8>.

- *Flexibility*: Hadoop can absorb any type of data, structured or not, from any type of source (geeks would say such a system is *schema-less*). But this disparate data can still be aggregated and analyzed.

- *Scalability*: Hadoop systems can start on a single PC, but thousands of machines can eventually be combined to work together for storage and analysis.
- *Cost effectiveness*: Since the system is open source and can be started with low-end hardware, the technology is cheap by data-warehousing standards. Many vendors also offer Hadoop as a cloud service, allowing firms to avoid hardware costs altogether.
- *Fault tolerance*: One of the servers running your Hadoop cluster just crashed? No big deal. Hadoop is designed in such a way so that there will be no single point of failure. The system will continue to work, relying on the remaining hardware.

Financial giant Morgan Stanley is a big believer in Hadoop. One senior technology manager at the firm contrasts Hadoop with highly structured systems, saying that in the past, “IT asked the business what they want, creates a data structure and writes structured query language, sources the data, conforms it to the table and writes a structured query. Then you give it to them and they often say that is not what they wanted.” But with Hadoop overseeing a big pile of unstructured (or less structured) data, technical staff can now work with users to carve up and combine data in lots of different ways, or even set systems loose in the data to hunt for unexpected patterns (see the discussion of data mining later in this chapter). Morgan Stanley’s initial Hadoop experiments started with a handful of old servers that were about to be retired, but the company has steadily ramped up its efforts. Now by using Hadoop, the firm sees that it is able to analyze data on a far larger scale (“petabytes of data, which is unheard of in the traditional database world”) with potentially higher-impact results. The bank is looking at customers’ financial objectives and trying to come up with investment insights to help them invest appropriately, and it is seeking “Big Data” insights to help the firm more effectively manage risk. T. Groenfeldt, “Morgan Stanley takes on Big Data with Hadoop,” *Forbes*, May 30, 2012.

Figure 11.3 The Hadoop Logo



The project was named after a toy elephant belonging to the son of Hadoop Developer Doug Cutting.

Other big name-firms using Hadoop for “Big Data” insights include Bank of America, Disney, GE, LinkedIn, Nokia, Twitter, and Wal-Mart. Hadoop is an open source project overseen by the Apache Software Foundation. It has an Internet pedigree and is based on ideas by Google and lots of software contributed by Yahoo! (two firms that regularly need to dive into massive and growing amounts of unstructured data—web pages, videos, images, social media, user account information, and more). IBM used Hadoop as the engine that helped power Watson to defeat human opponents on *Jeopardy*, further demonstrating the technology’s ability to analyze wildly different data for accurate insight. Other tech firms embracing Hadoop and offering some degree of support for the technology include HP, EMC, and Microsoft.

E-discovery: Supporting Legal Inquiries

Data archiving isn’t just for analytics. Sometimes the law requires organizations to dive into their electronic records. **E-discovery**²⁴ refers to identifying and retrieving relevant electronic information to support litigation efforts. E-discovery is something a firm should account for in its archiving and data storage plans. Unlike analytics that promise a boost to the bottom line, there’s no profit in complying with a judge’s order—it’s just a sunk cost. But organizations can be compelled by court order to scavenge their bits, and the cost to uncover difficult to access data can be significant, if not planned for in advance.

In one recent example, the Office of Federal Housing Enterprise Oversight (OFHEO) was subpoenaed for documents in litigation involving mortgage firms Fannie Mae and Freddie Mac. Even though the OFHEO wasn’t a party in the lawsuit, the agency had to comply with the search—an effort that cost \$6 million, a full 9 percent of its total yearly budget. A. Conry-Murray, “The Pain of E-discovery,” *InformationWeek*, June 1, 2009.

24. The process of identifying and retrieving relevant electronic information to support litigation efforts.

KEY TAKEAWAYS

- Data warehouses and data marts are repositories for large amounts of transactional data awaiting analytics and reporting.
- Large data warehouses are complex, can cost millions, and take years to build.
- The open source Hadoop effort provides a collection of technologies for manipulating massive amounts of unstructured data. The system is flexible, saleable, cost-effective, and fault-tolerant. Hadoop grew from large Internet firms but is now being used across industries.

QUESTIONS AND EXERCISES

1. List the issues that need to be addressed in order to design, develop, deploy, and maintain data warehouses and data marts.
2. What is meant by “data relevance”?
3. What is meant by “data governance”?
4. What is the difference between a data mart and a data warehouse?
5. Why are data marts and data warehouses necessary? Why can’t an organization simply query its transactional database?
6. How can something as simple as customer gender be difficult to for a large organization to establish in a data warehouse?
7. What is Hadoop? Why would a firm use Hadoop instead of conventional data warehousing and data mart technologies?
8. Research the current state of “Big Data” analysis. Identify major firms leveraging Hadoop or similar technologies and share high-impact examples with your classmates and professor.

11.6 The Business Intelligence Toolkit

LEARNING OBJECTIVES

1. Know the tools that are available to turn data into information.
2. Identify the key areas where businesses leverage data mining.
3. Understand some of the conditions under which analytical models can fail.
4. Recognize major categories of artificial intelligence and understand how organizations are leveraging this technology.

So far we've discussed where data can come from, and how we can get data into a form where we can use it. But how, exactly, do firms turn that data into information? That's where the various software tools of business intelligence (BI) and analytics come in. Potential products in the business intelligence toolkit range from simple spreadsheets to ultrasophisticated data mining packages leveraged by teams employing "rocket-science" mathematics.

Query and Reporting Tools

The idea behind query and reporting tools is to present users with a subset of requested data, selected, sorted, ordered, calculated, and compared, as needed. Managers use these tools to see and explore what's happening inside their organizations.

Canned reports²⁵ provide regular summaries of information in a predetermined format. They're often developed by information systems staff and formats can be difficult to alter. By contrast, **ad hoc reporting tools**²⁶ allow users to dive in and create their own reports, selecting fields, ranges, and other parameters to build their own reports on the fly. **Dashboards**²⁷ provide a sort of heads-up display of critical indicators, letting managers get a graphical glance at key performance metrics. Some tools may allow data to be exported into spreadsheets. Yes, even the lowly spreadsheet can be a powerful tool for modeling "what if" scenarios and creating additional reports (of course be careful: if data can be easily exported, then it can potentially leave the firm dangerously exposed, raising privacy, security, legal, and competitive concerns).

25. Reports that provide regular summaries of information in a predetermined format.

26. Tools that put users in control so that they can create custom reports on an as-needed basis by selecting fields, ranges, summary conditions, and other parameters.

27. A heads-up display of critical indicators that allow managers to get a graphical glance at key performance metrics.

Figure 11.4 *The Federal IT Dashboard*



The Federal IT dashboard offers federal agencies, and the general public, information about the government’s IT investments.

A subcategory of reporting tools is referred to as **online analytical processing (OLAP)**²⁸ (pronounced “oh-lap”). Data used in OLAP reporting is usually sourced from standard relational databases, but it’s calculated and summarized in advance, across multiple dimensions, with the data stored in a special database called a **data cube**²⁹. This extra setup step makes OLAP fast (sometimes one thousand times faster than performing comparable queries against conventional relational databases). Given this kind of speed boost, it’s not surprising that data cubes for OLAP access are often part of a firm’s data mart and data warehouse efforts.

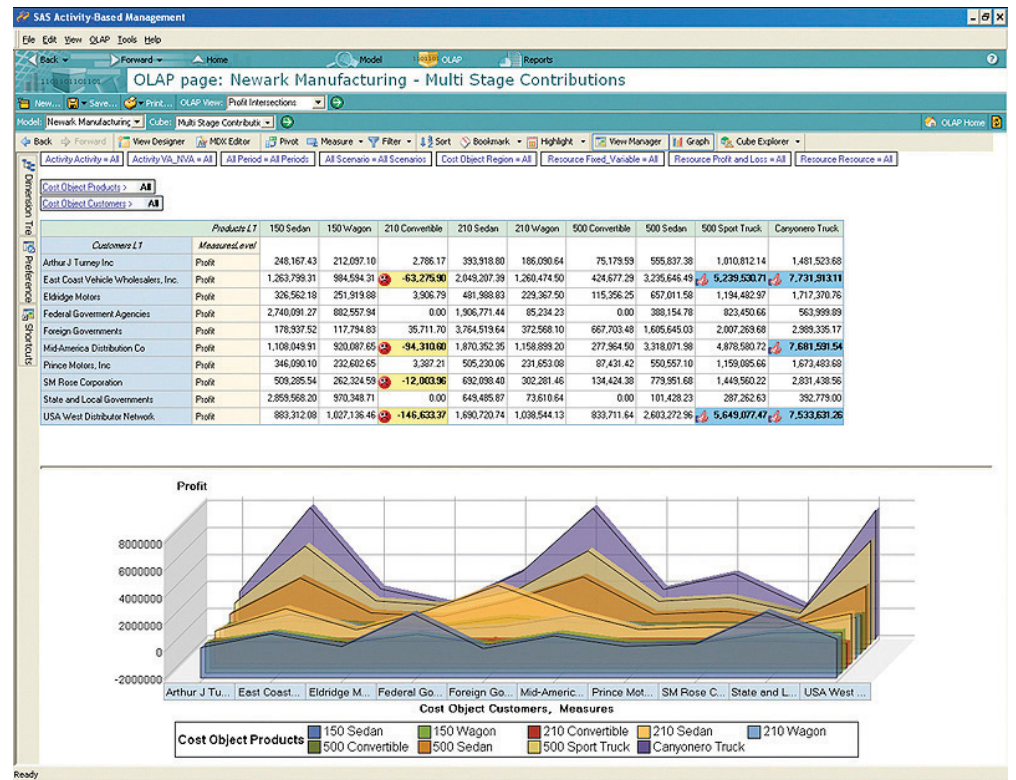
28. A method of querying and reporting that takes data from standard relational databases, calculates and summarizes the data, and then stores the data in a special database called a data cube.

29. A special database used to store data in OLAP reporting.

A manager using an OLAP tool can quickly explore and compare data across multiple factors such as time, geography, product lines, and so on. In fact, OLAP users often talk about how they can “slice and dice” their data, “drilling down” inside the data to uncover new insights. And while conventional reports are usually presented as a summarized list of information, OLAP results look more like a

spreadsheet, with the various dimensions of analysis in rows and columns, with summary values at the intersection.

Figure 11.5



This OLAP report compares multiple dimensions. Company is along the vertical axis, and product is along the horizontal access. Many OLAP tools can also present graphs of multidimensional data.

Copyright © 2009 SAS Institute, Inc.

Public Sector Reporting Tools in Action: Fighting Crime and Fighting Waste

Access to ad hoc query and reporting tools can empower all sorts of workers. Consider what analytics tools have done for the police force in Richmond, Virginia. The city provides department investigators with access to data from internal sources such as 911 logs and police reports, and combines this with outside data including neighborhood demographics, payday schedules, weather reports, traffic patterns, sports events, and more.

Experienced officers dive into this data, exploring when and where crimes occur. These insights help the department decide how to allocate its limited policing assets to achieve the biggest impact. While IT staffers put the system together, the tools are actually used by officers with expertise in fighting street crime—the kinds of users with the knowledge to hunt down trends and interpret the causes behind the data. And it seems this data helps make smart cops even smarter—the system is credited with delivering a single-year crime-rate reduction of 20 percent. S. Lohr, “Reaping Results: Data-Mining Goes Mainstream,” *New York Times*, May 20, 2007.

As it turns out, what works for cops also works for bureaucrats. When administrators for Albuquerque were given access to ad hoc reporting systems, they uncovered all sorts of anomalies, prompting excess spending cuts on everything from cell phone usage to unnecessarily scheduled overtime. And once again, BI performed for the public sector. The Albuquerque system delivered the equivalent of \$2 million in savings in just the first three weeks it was used. R. Mulcahy, “ABC: An Introduction to Business Intelligence,” *CIO*, March 6, 2007.

Data Mining

While reporting tools can help users explore data, modern data sets can be so large that it might be impossible for humans to spot underlying trends. That’s where data mining can help. **Data mining**³⁰ is the process of using computers to identify hidden patterns and to build models from large data sets.

30. The process of using computers to identify hidden patterns in, and to build models from, large data sets.

Some of the key areas where businesses are leveraging data mining include the following:

- *Customer segmentation*—figuring out which customers are likely to be the most valuable to a firm.
- *Marketing and promotion targeting*—identifying which customers will respond to which offers at which price at what time.
- *Market basket analysis*—determining which products customers buy together, and how an organization can use this information to cross-sell more products or services.
- *Collaborative filtering*—personalizing an individual customer’s experience based on the trends and preferences identified across similar customers.
- *Customer churn*—determining which customers are likely to leave, and what tactics can help the firm avoid unwanted defections.
- *Fraud detection*—uncovering patterns consistent with criminal activity.
- *Financial modeling*—building trading systems to capitalize on historical trends.
- *Hiring and promotion*—identifying characteristics consistent with employee success in the firm’s various roles.

For data mining to work, two critical conditions need to be present: (1) the organization must have clean, consistent data, and (2) the events in that data should reflect current and future trends. The recent financial crisis provides lessons on what can happen when either of these conditions isn’t met.

First let’s look at problems with using bad data. A report in the *New York Times* has suggested that in the period leading up to the 2008 financial crisis, some banking executives deliberately deceived risk management systems in order to skew capital-on-hand requirements. This deception let firms load up on risky debt, while carrying less cash for covering losses. S. Hansell, “How Wall Street Lied to Its Computers,” *New York Times*, September 18, 2008. Deceive your systems with bad data and your models are worthless. In this case, wrong estimates from bad data left firms grossly overexposed to risk. When debt defaults occurred; several banks failed, and we entered the worst financial crisis since the Great Depression.

Now consider the problem of historical consistency: Computer-driven investment models can be very effective when the market behaves as it has in the past. But models are blind when faced with the equivalent of the “hundred-year flood” (sometimes called *black swans*); events so extreme and unusual that they never showed up in the data used to build the model.

We saw this in the late 1990s with the collapse of the investment firm Long Term Capital Management. LTCM was started by Nobel Prize-winning economists, but when an unexpected Russian debt crisis caused the markets to move in ways not

anticipated by its models, the firm lost 90 percent of its value in less than two months. The problem was so bad that the Fed had to step in to supervise the firm's multibillion-dollar bailout. Fast forward a decade to the banking collapse of 2008, and we again see computer-driven trading funds plummet in the face of another unexpected event—the burst of the housing bubble. P. Wahba, “Buffeted ‘Quants’ Are Still in Demand,” *Reuters*, December 22, 2008.

Data mining presents a host of other perils, as well. It's possible to **over-engineer**³¹ a model, building it with so many variables that the solution arrived at might only work on the subset of data you've used to create it. You might also be looking at a random but meaningless statistical fluke. In demonstrating how flukes occur, one quantitative investment manager uncovered a correlation that at first glance appeared statistically to be a particularly strong predictor for historical prices in the S&P 500 stock index. That predictor? Butter production in Bangladesh. P. Coy, “He Who Mines Data May Strike Fool's Gold,” *BusinessWeek*, June 16, 1997. Sometimes durable and useful patterns just aren't in your data.

One way to test to see if you're looking at a random occurrence in the numbers is to divide your data, building your model with one portion of the data, and using another portion to verify your results. This is the approach Netflix has used to test results achieved by teams in the Netflix Prize, the firm's million-dollar contest for improving the predictive accuracy of its movie recommendation engine (see [Chapter 4 "Netflix in Two Acts: The Making of an E-commerce Giant and the Uncertain Future of Atoms to Bits"](#)).

Finally, sometimes a pattern is uncovered but determining the best choice for a response is less clear. As an example, let's return to the data-mining wizards at Tesco. An analysis of product sales data showed several money-losing products, including a type of bread known as “milk loaf.” Drop those products, right? Not so fast. Further analysis showed milk loaf was a “destination product” for a loyal group of high-value customers, and that these customers would shop elsewhere if milk loaf disappeared from Tesco shelves. The firm kept the bread as a loss-leader and retained those valuable milk loaf fans. B. Helm, “Getting Inside the Customer's Mind,” *BusinessWeek*, September 11, 2008. Data miner, beware—first findings don't always reveal an optimal course of action.

This last example underscores the importance of recruiting a data mining and business analytics team that possesses three critical skills: information technology (for understanding how to pull together data, and for selecting analysis tools), statistics (for building models and interpreting the strength and validity of results), and business knowledge (for helping set system goals, requirements, and offering deeper insight into what the data really says about the firm's operating

31. Build a model with so many variables that the solution arrived at might only work on the subset of data you've used to create it.

environment). Miss one of these key functions and your team could make some major mistakes.

While we've focused on tools in our discussion above, many experts suggest that business intelligence is really an organizational process as much as it is a set of technologies. Having the right team is critical in moving the firm from goal setting through execution and results.

Artificial Intelligence

Data mining has its roots in a branch of computer science known as artificial intelligence (or AI). The goal of AI is create computer programs that are able to mimic or improve upon functions of the human brain. Data mining can leverage **neural networks**³² or other advanced algorithms and statistical techniques to hunt down and expose patterns, and build models to exploit findings.

Expert systems³³ are AI systems that leverage rules or examples to perform a task in a way that mimics applied human expertise. Expert systems are used in tasks ranging from medical diagnoses to product configuration.

Genetic algorithms³⁴ are model building techniques where computers examine many potential solutions to a problem, iteratively modifying (mutating) various mathematical models, and comparing the mutated models to search for a best alternative. Genetic algorithms have been used to build everything from financial trading models to handling complex airport scheduling, to designing parts for the international space station. Adapted from J. Kahn, "It's Alive," *Wired*, March 2002; O. Port, "Thinking Machines," *BusinessWeek*, August 7, 2000; and L. McKay, "Decisions, Decisions," *CRM Magazine*, May 1, 2009.

While AI is not a single technology, and not directly related to data creation, various forms of AI can show up as part of analytics products, CRM tools, transaction processing systems, and other information systems.

32. An AI system that examines data and hunts down and exposes patterns, in order to build models to exploit findings.
33. AI systems that leverages rules or examples to perform a task in a way that mimics applied human expertise.
34. Model building techniques where computers examine many potential solutions to a problem, iteratively modifying (mutating) various mathematical models, and comparing the mutated models to search for a best alternative.

KEY TAKEAWAYS

- Canned and ad hoc reports, digital dashboards, and OLAP are all used to transform data into information.
- OLAP reporting leverage data cubes, which take data from standard relational databases, calculating and summarizing data for superfast reporting access. OLAP tools can present results through multidimensional graphs, or via spreadsheet-style cross-tab reports.
- Modern data sets can be so large that it might be impossible for humans to spot underlying trends without the use of data mining tools.
- Businesses are using data mining to address issues in several key areas including customer segmentation, marketing and promotion targeting, collaborative filtering, and so on.
- Models influenced by bad data, missing or incomplete historical data, and over-engineering are prone to yield bad results.
- One way to test to see if you're looking at a random occurrence in your data is to divide your data, building your model with one portion of the data, and using another portion to verify your results.
- Analytics may not always provide the total solution for a problem. Sometimes a pattern is uncovered, but determining the best choice for a response is less clear.
- A competent business analytics team should possess three critical skills: information technology, statistics, and business knowledge.

QUESTIONS AND EXERCISES

1. What are some of the tools used to convert data into information?
2. What is the difference between a canned reports and an ad hoc reporting?
3. How do reports created by OLAP differ from most conventional reports?
4. List the key areas where businesses are leveraging data mining.
5. What is market basket analysis?
6. What is customer churn?
7. For data mining to work, what two critical data-related conditions must be present?
8. Discuss occurrences of model failure caused by missing or incomplete historical data.
9. Discuss Tesco's response to their discovery that "milk loaf" was a money-losing product.
10. List the three critical skills a competent business analytics team should possess.
11. Do any of the products that you use leverage artificial intelligence? What kinds of AI might be used in Netflix's movie recommendation system, Apple's iTunes Genius playlist builder, or Amazon's Web site personalization? What kind of AI might help a physician make a diagnosis or help an engineer configure a complicated product in the field?

11.7 Data Asset in Action: Technology and the Rise of Wal-Mart

LEARNING OBJECTIVES

1. Understand how Wal-Mart has leveraged information technology to become the world's largest retailer.
2. Be aware of the challenges that face Wal-Mart in the years ahead.

Wal-Mart demonstrates how a physical product retailer can create and leverage a data asset to achieve world-class supply chain efficiencies targeted primarily at driving down costs.

Wal-Mart isn't just the largest retailer in the world, over the past several years it has popped in and out of the top spot on the *Fortune* 500 list—meaning that the firm has had revenues greater than *any* firm in the United States. Wal-Mart is so big that in three months it sells more than a whole year's worth of sales at number two U.S. retailer, Home Depot. From 2006 through 2009, Wal-Mart has appeared as either number one or number two in the *Fortune* 100 rankings.

At that size, it's clear that Wal-Mart's key source of competitive advantage is scale. But firms don't turn into giants overnight. Wal-Mart grew in large part by leveraging information systems to an extent never before seen in the retail industry. Technology tightly coordinates the Wal-Mart value chain from tip to tail, while these systems also deliver a mineable data asset that's unmatched in U.S. retail. To get a sense of the firm's overall efficiencies, at the end of the prior decade a McKinsey study found that Wal-Mart was responsible for some 12 percent of the productivity gains in the *entire* U.S. economy. C. Fishman, "The Wal-Mart You Don't Know," *Fast Company*, December 19, 2007. The firm's capacity as a systems innovator is so respected that many senior Wal-Mart IT executives have been snatched up for top roles at Dell, HP, Amazon, and Microsoft. And lest one think that innovation is the province of only those located in the technology hubs of Silicon Valley, Boston, and Seattle, remember that Wal-Mart is headquartered in Bentonville, Arkansas.

A Data-Driven Value Chain

The Wal-Mart efficiency dance starts with a proprietary system called Retail Link, a system originally developed in 1991 and continually refined ever since. Each time an item is scanned by a Wal-Mart cash register, Retail Link not only records the sale, it also automatically triggers inventory reordering, scheduling, and delivery.

This process keeps shelves stocked, while keeping inventories at a minimum. An AMR report ranked Wal-Mart as having the seventh best supply chain in the country (the only other retailer in the top twenty was Tesco, at number fifteen). T. Friscia, K. O'Marah, D. Hofman, and J. Souza, "The AMR Research Supply Chain Top 25 for 2009," *AMR Research*, May 28, 2009, <http://www.amrresearch.com/Content/View.aspx?compURI=tcm:7-43469>. The firm's annual **inventory turnover ratio**³⁵ of 8.5 means that Wal-Mart sells the equivalent of its entire inventory roughly every six weeks (by comparison, Target's turnover ratio is 6.4, Sears' is 3.4, and the average for U.S. retail is less than 2). Twelve-month figures from midyear 2009, via *Forbes* and Reuters.

Back-office scanners keep track of inventory as supplier shipments come in. Suppliers are rated based on timeliness of deliveries, and you've got to be quick to work with Wal-Mart. In order to avoid a tractor-trailer traffic jam in store parking lots, deliveries are choreographed to arrive at intervals less than ten minutes apart. When Levi's joined Wal-Mart, the firm had to guarantee it could replenish shelves every two days—no prior retailer had required a shorter than five day window from Levi's. C. Fishman, "The Wal-Mart You Don't Know," *Fast Company*, December 19, 2007.

Wal-Mart has been a catalyst for technology adoption among its suppliers. The firm is currently leading an adoption effort that requires partners to leverage RFID technology to track and coordinate inventories. While the rollout has been slow, a recent P&G trial showed RFID boosted sales nearly 20 percent by ensuring that inventory was on shelves and located where it should be. D. Joseph, "Supermarket Strategies: What's New at the Grocer," *BusinessWeek*, June 8, 2009.

Data Mining Prowess

Wal-Mart also mines its mother lode of data to get its product mix right under all sorts of varying environmental conditions, protecting the firm from "a retailer's twin nightmares: too much inventory, or not enough." C. Hays, "What Wal-Mart Knows about Customer Habits," *New York Times*, November 14, 2004. For example, the firm's data mining efforts informed buyers that customers stock up on certain products in the days leading up to predicted hurricanes. Bumping up prestorm supplies of batteries and bottled water was a no brainer, but the firm also learned that Pop-Tarts sales spike sevenfold before storms hit, and that beer is the top prestorm seller. This insight has led to truckloads full of six packs and toaster pastries streaming into gulf states whenever word of a big storm surfaces. C. Hays, "What Wal-Mart Knows about Customer Habits," *New York Times*, November 14, 2004.

35. The ratio of a company's annual sales to its inventory.

Data mining also helps the firm tighten operational forecasts, helping to predict things like how many cashiers are needed at a given store at various times of day throughout the year. Data drives the organization, with mined reports forming the basis of weekly sales meetings, as well as executive strategy sessions.

Wal-Mart leverages its huge Hadoop-based data trove to support some of its data mining efforts, sifting through massive amounts of social media—Twitter posts, Facebook updates, and other so-called unstructured data—to gain insights on product offerings, sales leads, pricing, and more. The firm purchased social startup Kosmix for \$300 million to deepen social and Big Data expertise in the company's @WalmartLabs.R. King, "Getting a Handle on Big Data with Hadoop," *BusinessWeek*, September 7, 2011. Says Kosmix founder Anand Rajaraman (who previously sold a firm to Amazon and was an early investor in Facebook), "The first generation of e-commerce was about bringing the store to the Web. The next generation will be about building integrated experiences that leverage the store, the Web and mobile, with social identity being the glue that binds the experience."C. Nicholson, "Wal-Mart Buys Social Media Firm Kosmix," *New York Times*, April 19, 2011.

Sharing Data, Keeping Secrets

While Wal-Mart is demanding of its suppliers, it also shares data with them, too. Data can help firms become more efficient so that Wal-Mart can keep dropping prices, and data can help firms uncover patterns that help suppliers sell more. P&G's Gillette unit, for example, claims to have mined Wal-Mart data to develop promotions that increased sales as much as 19 percent. More than seventeen thousand suppliers are given access to their products' Wal-Mart performance across metrics that include daily sales, shipments, returns, purchase orders, invoices, claims and forecasts. And these suppliers collectively interrogate Wal-Mart data warehouses to the tune of twenty-one million queries a year.K. Evans-Correia, "Dillman Replaced as Wal-Mart CIO," *SearchCIO*, April 6, 2006.

While Wal-Mart shares sales data with relevant suppliers, the firm otherwise fiercely guards this asset. Many retailers pool their data by sharing it with information brokers like Information Resources and ACNielsen. This sharing allows smaller firms to pool their data to provide more comprehensive insight on market behavior. But Wal-Mart stopped sharing data with these agencies years ago. The firm's scale is so big, the additional data provided by brokers wasn't adding much value, and it no longer made sense to allow competitors access to what was happening in its own huge chunk of retail sales.

Other aspects of the firm's technology remain under wraps, too. Wal-Mart custom builds large portions of its information systems to keep competitors off its trail. As

for infrastructure secrets, the Wal-Mart Data Center in McDonald County, Missouri, was considered so off limits that the county assessor was required to sign a nondisclosure statement before being allowed on-site to estimate property value. M. McCoy, "Wal-Mart's Data Center Remains Mystery," *Joplin Globe*, May 28, 2006.

Challenges Abound

But despite success, challenges continue. While Wal-Mart grew dramatically throughout the 1990s, the firm's U.S. business has largely matured. And as a mature business it faces a problem not unlike the example of Microsoft discussed at the end of [Chapter 14 "Google in Three Parts: Search, Online Advertising, and Beyond"](#); Wal-Mart needs to find huge markets or dramatic cost savings in order to boost profits and continue to move its stock price higher.

The firm's aggressiveness and sheer size also increasingly make Wal-Mart a target for criticism. Those low prices come at a price, and the firm has faced accusations of subpar wages and remains a magnet for union activists. Others had identified poor labor conditions at some of the firm's contract manufacturers. Suppliers that compete for Wal-Mart's business are often faced with a catch-22. If they bypass Wal-Mart they miss out on the largest single chunk of world retail sales. But if they sell to Wal-Mart, the firm may demand prices so aggressively low that suppliers end up cannibalizing their own sales at other retailers. Still more criticism comes from local citizen groups that have accused Wal-Mart of ruining the market for mom-and-pop stores. C. Fishman, "The Wal-Mart You Don't Know," *Fast Company*, December 19, 2007.

While some might see Wal-Mart as invincibly standing at the summit of world retail, it's important to note that other megaretailers have fallen from grace. In the 1920s and 1930s, the A&P grocery chain once controlled 80 percent of U.S. grocery sales, at its peak operating five times the number of stores that Wal-Mart has today. But market conditions changed, and the government stepped in to draft antipredatory pricing laws when it felt A&P's parent was too aggressive.

For all of Wal-Mart's data brilliance, historical data offers little insight on how to adapt to more radical changes in the retail landscape. The firm's data warehouse wasn't able to foretell the rise of Target and other up-market discounters. And yet another major battle is brewing, as Tesco methodically attempts to take its globally honed expertise to U.S. shores. Savvy managers recognize that data use is a vital tool, but not the only tool in management's strategic arsenal.

KEY TAKEAWAYS

- Wal-Mart demonstrates how a physical product retailer can create and leverage a data asset to achieve world-class value chain efficiencies.
- Wal-Mart uses data mining in numerous ways, from demand forecasting to predicting the number of cashiers needed at a store at a particular time.
- To help suppliers become more efficient, and as a result lower prices, Wal-Mart shares data with them.
- Despite its success, Wal-Mart is a mature business that needs to find huge markets or dramatic cost savings in order to boost profits and continue to move its stock price higher. The firm's success also makes it a high impact target for criticism and activism. And the firm's data assets could not predict impactful industry trends such as the rise of Target and other upscale discounters.

QUESTIONS AND EXERCISES

1. List the functions performed by Retail Link. What is its benefit to Wal-Mart?
2. Which supplier metrics does Retail Link gather and report? How is this valuable to Wal-Mart and suppliers?
3. Name the technology does Wal-Mart require partners to use to track and coordinate inventory. Do you know of other uses for this technology?
4. What steps has Wal-Mart taken to protect its data from competitors?
5. List the criticisms leveled at Wal-Mart. Do you think these critiques are valid or not? What can Wal-Mart do to counteract this criticism? Should it take these steps? Why or why not?

11.8 Data Asset in Action: Caesars' Solid Gold CRM for the Service Sector

LEARNING OBJECTIVES

1. Understand how Caesars has used IT to move from an also-ran chain of casinos to become the largest gaming company based on revenue.
2. Name some of the technology innovations that Caesars is using to help it gather more data, and help push service quality and marketing program success.

Caesars Entertainment (formerly known by the name of its acquirer, Harrah's) provides an example of exceptional data asset leverage in the service sector, focusing on how this technology enables world-class service through customer relationship management. And as you read this case, keep in mind that the firm's CEO, Gary Loveman, claims that what he did at Caesars he could have done at most firms in most other industries. D. Talbot, "Using IT to Drive Innovation," *Technology Review*, February 16, 2011.

Gary Loveman is a sort of management major trifecta. The CEO of Caesars Entertainment is a former *operations* professor who has leveraged *information technology* to create what may be the most effective *marketing* organization in the service industry. If you ever needed an incentive to motivate you for cross-disciplinary thinking, Loveman provides it.

Caesars has leveraged its data-powered prowess to move from an also-ran chain of casinos to become the largest gaming company by revenue. The firm operates some fifty-three casinos, employing more than eighty-five thousand workers on five continents. Brands include Harrah's, Caesars Palace, Bally's, Horseshoe, and Paris Las Vegas. Under Loveman's leadership, the firm formerly known as Harrah's aggressively swallowed competitors, with the firm's \$9.4 billion buyout of Caesars Entertainment being its largest deal to date (while as separate firms, Harrah's under Loveman trounced Caesars in financial performance, the Caesars name was seen as a stronger brand).

Collecting Data

Data drives the firm. Caesars collects customer data on just about everything you might do at their properties—gamble, eat, grab a drink, attend a show, stay in a room. The data's then used to track your preferences and to size up whether you're

the kind of customer that's worth pursuing. Prove your worth, and the firm will surround you with top-tier service and develop a targeted marketing campaign to keep wooing you back. V. Magnini, E. Honeycutt, and S. Hodge, "Data Mining for Hotel Firms: Use and Limitations," *Cornell Hotel and Restaurant Administration Quarterly*, April 2003, <http://www.entrepreneur.com/tradejournals/article/101938457.html>.

The ace in the firm's data collection hole is its Total Rewards loyalty card system. Launched over a decade ago, the system is constantly being enhanced by an IT staff of seven hundred, with an annual budget in excess of \$100 million. P. Swabey, "Nothing Left to Chance," *Information Age*, January 18, 2007. Total Rewards is an **opt-in**³⁶ loyalty program, but customers consider the incentives to be so good that the card is used by some 80 percent of patrons, collecting data on over forty-four million customers. M. Wagner, "Harrah's Places Its Bet On IT," *InformationWeek*, September 16, 2008; and L. Haugsted, "Better Take Care of Big Spenders; Harrah's Chief Offers Advice to Cablers," *Multichannel News*, July 30, 2007.

Customers signing up for the card provide Caesars with demographic information such as gender, age, and address. Visitors then present the card for various transactions. Slide it into a slot machine, show it to the restaurant hostess, present it to the parking valet, share your account number with a telephone reservation specialist—every contact point is an opportunity to collect data. Between three hundred thousand and one million customers come through Caesars' doors daily, adding to the firm's data stash and keeping that asset fresh. N. Hoover, "Chief of the Year: Harrah's CIO Tim Stanley," *Information Week Research and Reports*, 2007.

Who Are the Most Valuable Customers?

All that data is heavily and relentlessly mined. Customer relationship management should include an assessment to determine which customers are worth having a relationship with. And because Caesars has so much detailed historical data, the firm can make fairly accurate projections of **customer lifetime value (CLV)**³⁷. CLV represents the present value of the likely future income stream generated by an individual purchaser. "Which Customers Are Worth Keeping and Which Ones Aren't? Managerial Uses of CLV," *Knowledge@Wharton*, July 30, 2003, <http://knowledge.wharton.upenn.edu/article.cfm?articleid=820>. Once you know this, you can get a sense of how much you should spend to keep that customer coming back. You can size them up next to their peer group, and if they fall below expectations, you can develop strategies to improve their spending.

The firm tracks over ninety demographic segments, and each responds differently to different marketing approaches. Identifying segments and figuring out how to

36. Program (typically a marketing effort) that requires customer consent. This program is contrasted with opt-out programs, which enroll all customers by default.

37. The present value of the likely future income stream generated by an individual purchaser.

deal with each involves an iterative model of mining the data to identify patterns, creating a hypothesis (customers in group X will respond to a free steak dinner; group Y will want ten dollars in casino chips), then testing that hypothesis against a control group, turning again to analytics to statistically verify the outcome.

The firm runs hundreds of these small, controlled experiments each year. Loveman says that when marketers suggest new initiatives, “I ask, did we test it first? And if I find out that we just whole-hogged, went after something without testing it, I’ll kill ’em. No matter how clever they think it is, we test it.” J. Nickell, “Welcome to Harrah’s,” *Business 2.0*, April 2002. The former ops professor is known to often quote quality guru W. Edwards Deming, saying, “In God we trust; all others must bring data.”

When Caesars began diving into the data, they uncovered patterns that defied the conventional wisdom in the gaming industry. Big money didn’t come from European princes, Hong Kong shipping heirs, or the *Ocean’s 11* crowd—it came from locals. The less than 30 percent of customers who spent between one hundred and five hundred dollars per visit accounted for over 80 percent of revenues and nearly 100 percent of profits. P. Swabey, “Nothing Left to Chance,” *Information Age*, January 18, 2007.

The data also showed that the firm’s most important customers weren’t the families that many Vegas competitors were trying to woo with Disneyland-style theme casinos—it was Grandma! The firm focuses on customers forty-five years and older: twenty-somethings have no money, while thirty-somethings have kids and are too busy. To the premiddle-aged crowd, Loveman says, “God bless you, but we don’t need you.” L. Haugsted, “Better Take Care of Big Spenders; Harrah’s Chief Offers Advice to Cablers,” *Multichannel News*, July 30, 2007.

Data-Driven Service: Get Close (but Not Too Close) to Your Customers

The names for reward levels on the Total Rewards card convey increasing customer value—Gold, Diamond, and Platinum. Spend more money at Caesars and you’ll enjoy shorter lines, discounts, free items, and more. And if Caesars’ systems determine you’re a high-value customer, expect white-glove treatment. The firm will lavish you with attention, using technology to try to anticipate your every need. Customers notice the extra treatment that top-tier Total Rewards members receive and actively work to improve their status.

To illustrate this, Loveman points to the obituary of an Asheville, North Carolina, woman who frequented a casino his firm operates on a nearby Cherokee reservation. “Her obituary was published in the Asheville paper and indicated that at the time of her death, she had several grandchildren, she sang in the Baptist choir and she was *a holder of the [the firm’s] Diamond Total Rewards card.*” Quipped Loveman, “When your loyalty card is listed in someone’s obituary, I would maintain *you have traction.*”G. Loveman, Speech and Comments, Chief Executive Club of Boston College, January 2005; emphasis added.

The degree of customer service pushed through the system is astonishing. Upon check-in, a Caesars customer who enjoys fine dining may find his or her table is reserved, along with tickets for a show afterward. Others may get suggestions or special offers throughout their stay, pushed via text message to their mobile device.M. Wagner, “Harrah’s Places Its Bet On IT,” *InformationWeek*, September 16, 2008. The firm even tracks gamblers to see if they’re suffering unusual losses, and Caesars will dispatch service people to intervene with a feel-good offer: “Having a bad day? Here’s a free buffet coupon.”T. Davenport and J. Harris, *Competing on Analytics: The New Science of Winning* (Boston: Harvard Business School Press, 2007).

The firm’s CRM effort monitors any customer behavior changes. If a customer who usually spends a few hundred a month hasn’t shown up in a while, the firm’s systems trigger follow-up contact methods such as sending a letter with a promotion offer, or having a rep make a phone call inviting them back.G. Loveman, Speech and Comments, Chief Executive Club of Boston College, January 2005.

Customers come back to Caesars because they feel that those casinos treat them better than the competition. And Caesars’ laser-like focus on service quality and customer satisfaction are embedded into its information systems and operational procedures. Employees are measured on metrics that include speed and friendliness and are compensated based on guest satisfaction ratings. Hourly workers are notoriously difficult to motivate: they tend to be high-turnover, low-wage earners. But at Caesars, incentive bonuses depend on an entire location’s ratings. That encourages strong performers to share tips to bring the new guy up to speed. The process effectively changed the corporate culture at Caesars from an every-property-for-itself mentality to a collaborative, customer-focused enterprise.V. Magnini, E. Honeycutt, and S. Hodge, “Data Mining for Hotel Firms: Use and Limitations,” *Cornell Hotel and Restaurant Administration Quarterly*, April 2003, <http://www.entrepreneur.com/tradejournals/article/101938457.html>.

While Caesars is committed to learning how to make your customer experience better, the firm is also keenly sensitive to respecting consumer data. The firm has never sold or given away any of its bits to third parties. And the firm admits that

some of its efforts to track customers have misfired, requiring special attention to find the sometimes subtle line between helpful and “too helpful.” For example, the firm’s CIO has mentioned that customers found it “creepy and Big Brother-ish” when employees tried to greet them by name and talk with them about their past business history with the firm, so it backed off. M. Wagner, “Harrah’s Places Its Bet On IT,” *InformationWeek*, September 16, 2008.

Innovation

Caesars is constantly tinkering with new innovations that help it gather more data and help push service quality and marketing program success. When the introduction of gaming in Pennsylvania threatened to divert lucrative New York City gamblers from Caesars’ Atlantic City properties, the firm launched an interactive billboard in New York’s Times Square, allowing passersby to operate a virtual slot machine using text messages from their cell phones. Players dialing into the video billboard not only control the display, they receive text message offers promoting Caesars’ sites in Atlantic City. “Future Tense: The Global CMO,” *Economist Intelligence Unit*, September 2008.

At Caesars, tech experiments abound. RFID-enabled poker chips and under-table RFID readers allow pit bosses to track and rate game play far better than they could before. The firm is experimenting with using RFID-embedded bracelets for poolside purchases and Total Rewards tracking for when customers aren’t carrying their wallets. The firm has also incorporated drink ordering into gaming machines—why make customers get up to quench their thirst? A break in gambling is a halt in revenue.

The firm was also one of the first to sign on to use Microsoft’s Surface technology—a sort of touch-screen and sensor-equipped tabletop. Customers at these tables can play bowling and group pinball games and even pay for drinks using cards that the tables will automatically identify. Tech even helps Caesars fight card counters and crooks, with facial recognition software scanning casino patrons to spot the bad guys. S. Lohr, “Reaping Results: Data-Mining Goes Mainstream,” *New York Times*, May 20, 2007.

And Total Rewards is going social, too. Caesars has partnered with Silicon Valley-based TopGuest to tie social media to its loyalty program. Caesars customers who register with TopGuest (which also works with clients Virgin America, Holiday Inn, and Avis, among others) can receive fifty Total Rewards bonus credits for each geolocation check-in, tweet, or Instagram photo taken at participating venues. “Topguest Rewards Vegas Visitors for Social Check-Ins,” *Pulse of Vegas Blog* (hosted at Harrahs.com), April 22, 2011.

Strategy

A walk around Vegas during Caesars' ascendancy would find rivals with bigger, fancier casinos. Says Loveman, "We had to compete with the kind of place that God would build if he had the money....The only thing we had was data." P. Swabey, "Nothing Left to Chance," *Information Age*, January 18, 2007.

That data advantage creates intelligence for a high-quality and highly personal customer experience. Data gives the firm a service differentiation edge. The loyalty program also represents a switching cost. And these assets combined to be leveraged across a firm that has gained so much scale that it's now the largest player in its industry, gaining the ability to cross-sell customers on a variety of properties—Vegas vacations, riverboat gambling, locally focused reservation properties, and more.

The firm's chief marketing officer points out that when the Total Rewards effort started, the firm was earning about thirty-six cents on every dollar customers spent gaming—the rest went to competitors. A climb to forty cents would be considered monstrous. But within a few short years that number had climbed to forty-five cents, making Caesars the biggest monster in the industry. E. Lundquist, "Harrah's Bets Big on IT," *eWeek*, July 20, 2005. Some of the firm's technology investments have paid back tenfold in just two years—bringing in hundreds of millions of dollars. P. Swabey, "Nothing Left to Chance," *Information Age*, January 18, 2007.

The firm's technology has been pretty tough for others to match, too. Caesars holds several patents covering key business methods and technologies used in its systems. After being acquired by Harrah's, employees of the old Caesars properties lamented that they had, for years, unsuccessfully attempted to replicate Harrah's systems without violating the firm's intellectual property. N. Hoover, "Chief of the Year: Harrah's CIO Tim Stanley," *Information Week Research and Reports*, 2007.

Challenges

Caesars' efforts to gather data, extract information, and turn this into real profits is unparalleled, but it's not a cure-all. Broader events can often derail even the best strategy. Gaming is a discretionary spending item, and when the economy tanks, gambling is one of the first things consumers will cut. Caesars has not been immune to the world financial crisis and experienced a loss in 2008.

Also note that if you look up Caesars' stock symbol you won't find it. The firm was **taken private**³⁸ in January 2008, when buyout firms Apollo Management and TPG Capital paid \$30.7 billion for all of the firm's shares. At that time Loveman signed a

38. The process by which a publicly held company has its outstanding shares purchased by an individual or by a small group of individuals who wish to obtain complete ownership and control.

five-year deal to remain on as CEO, and he's spoken positively about the benefits of being private—primarily that with the distraction of quarterly earnings off the table, he's been able to focus on the long-term viability and health of the business.A. Knightly, "Harrah's Boss Speaks," *Las Vegas Review-Journal*, June 14, 2009. Plans for a late 2010 IPO were put on hold as the economy continued to sour.C. Vannucci and L. Spears, "Harrah's Pulls \$531 Million Private Equity-Backed IPO," *BusinessWeek*, November 19, 2011.

But the firm also holds \$24 billion in debt from expansion projects and the buyout, all at a time when economic conditions have not been favorable to leveraged firms.P. Lattman, "A Buyout-Shop Breather," *Wall Street Journal*, May 30, 2009. A brilliantly successful firm that developed best-in-class customer relationship management is now in a position many consider risky due to debt assumed as part of an overly optimistic buyout occurring at precisely the time when the economy went into a terrible funk. Caesars' awesome risk-reducing, profit-pushing analytics failed to offer any insight on the wisdom (or risk) in the debt and private equity deals.

KEY TAKEAWAYS

- Caesars Entertainment provides an example of exceptional data asset leverage in the service sector, focusing on how this technology enables world-class service through customer relationship management.
- Caesars uses its Total Rewards loyalty card system to collect customer data on just about everything you might do at their properties—gamble, eat, drink, see a show, stay in a room, and so on.
- Individual customers signing up for the Total Rewards loyalty card provide Caesars with demographic information such as gender, age, and address, which is combined with transactional data as the card is used.
- Data mining also provides information about ninety-plus customer demographic segments, each of which responds differently to different marketing approaches.
- If Caesars' systems determine you're a high-value customer, you can expect a higher level of perks and service.
- Caesars' CRM effort monitors any customer behavior changes.
- Caesars uses its information systems and operating procedures to measure employees based on metrics that include speed and friendliness and compensates them based on guest satisfaction ratings.

QUESTIONS AND EXERCISES

1. What types of customer data does Caesars gather?
2. How is the data that Caesars collects used?
3. Describe Caesars' most valuable customers. Approximately what percentage of profits does this broad group deliver to the firm?
4. List the services a Rewards Card cardholder might expect.
5. What happens when a good, regular customer stops showing up?
6. Describe how Caesars treats customer data.
7. List some of the technology innovations that Caesars is using to help it gather more data and help push service quality and marketing program success.
8. How does Caesars' Total Rewards loyalty card system represent a switching cost?
9. What is customer lifetime value? Do you think this is an easier metric to calculate at Caesars or Wal-Mart? Why?
10. How did intellectual property protection benefit Caesars?
11. Discuss the challenges Caesars may have to confront in the near future.
12. Describe the role that testing plays in initiatives? What advantage does testing provide the firm? What's the CEO's attitude to testing? Do you agree with this level of commitment? Why or why not?
13. Do you think the firm's described foray into social media is a good idea or a bad one? Why? Discuss any potential upsides and downsides that you see coming from the effort.