

BUSINESS DRIVEN ETHICS AND SECURITY

Mining large amounts of data can create a number of benefits for business, society, and governments, but it can also create a number of ethical questions surrounding an invasion of privacy or misuse of information. Facebook recently came under fire for its data mining practices as it followed 700,000 accounts to determine whether posts with highly emotional content are more contagious. The study concluded that highly emotional texts are contagious, just as with real people. Highly emotional positive posts received multiple positive replies whereas highly emotional negative posts received multiple negative replies. Although the study seems rather innocent, many Facebook users were outraged; they felt the study was an invasion of privacy because the 700,000 accounts had no idea Facebook was mining their posts. As a Facebook user, you willingly consent that Facebook owns every bit and byte of data you post and, once you press submit, Facebook can do whatever it wants with your data. Do you agree or disagree that Facebook has the right to do whatever it wants with the data its 1.5 billion users post on its site?⁵

Unethical Data Mining

DRIVING WEBSITES WITH DATA

A **content creator** is the person responsible for creating the original website content. A **content editor** is the person responsible for updating and maintaining website content. **Static information** includes fixed data incapable of change in the event of a user action. **Dynamic information** includes data that change based on user actions. For example, static websites supply only information that will not change until the content editor changes the information. Dynamic information changes when a user requests information. A dynamic website changes information based on user requests such as movie ticket availability, airline prices, or restaurant reservations. Dynamic website information is stored in a **dynamic catalog**, or an area of a website that stores information about products in a database.

Websites change for site visitors depending on the type of information they request. Consider, for example, an automobile dealer. The dealer would create a database containing data elements for each car it has available for sale, including make, model, color, year, miles per gallon, a photograph, and so on. Website visitors might click Porsche and then enter their specific requests such as price range or year made. Once the user hits Go, the website automatically provides a custom view of the requested information. The dealer must create, update, and delete automobile information as the inventory changes.

A **data-driven website** is an interactive website kept constantly updated and relevant to the needs of its customers using a database. Data-driven capabilities are especially useful when a firm needs to offer large amounts of information, products, or services. Visitors can become quickly annoyed if they find themselves buried under an avalanche of information when searching a website. A data-driven website can help limit the amount of information displayed to customers based on unique search requirements. Companies even use data-driven websites to make information in their internal databases available to customers and business partners.

There are a number of advantages to using the web to access company databases. First, web browsers are much easier to use than directly accessing the database by using a custom-query tool. Second, the web interface requires few or no changes to the database model. Finally, it costs less to add a web interface in front of a DBMS than to redesign and rebuild the system to support changes. Additional data-driven website advantages include:

- ✱ **Easy to manage content:** Website owners can make changes without relying on MIS professionals; users can update a data-driven website with little or no training.

LO 6.4: Explain the business benefits of a data-driven website.

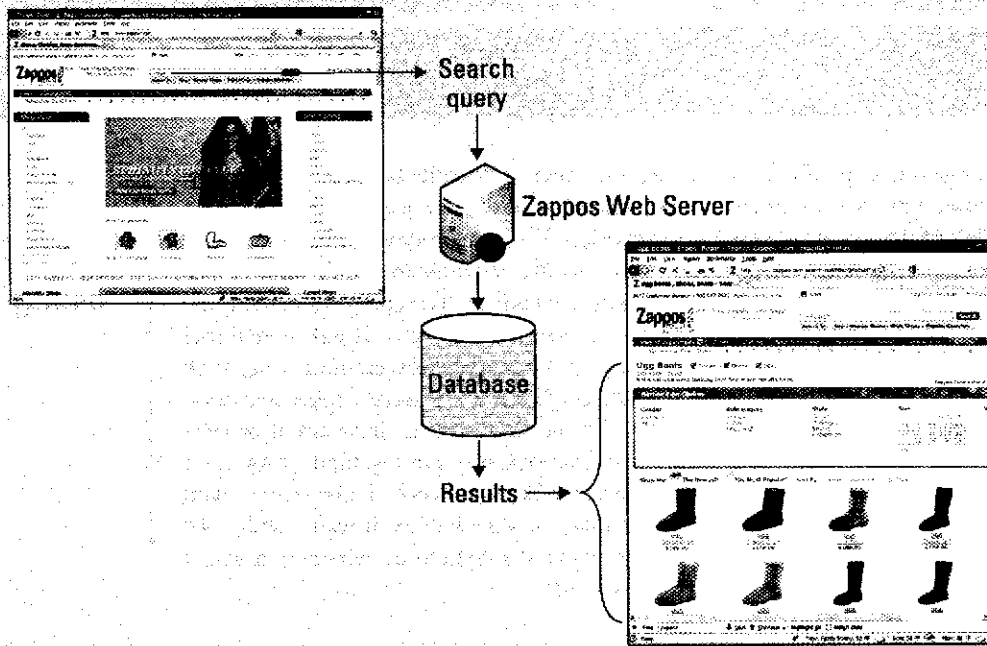
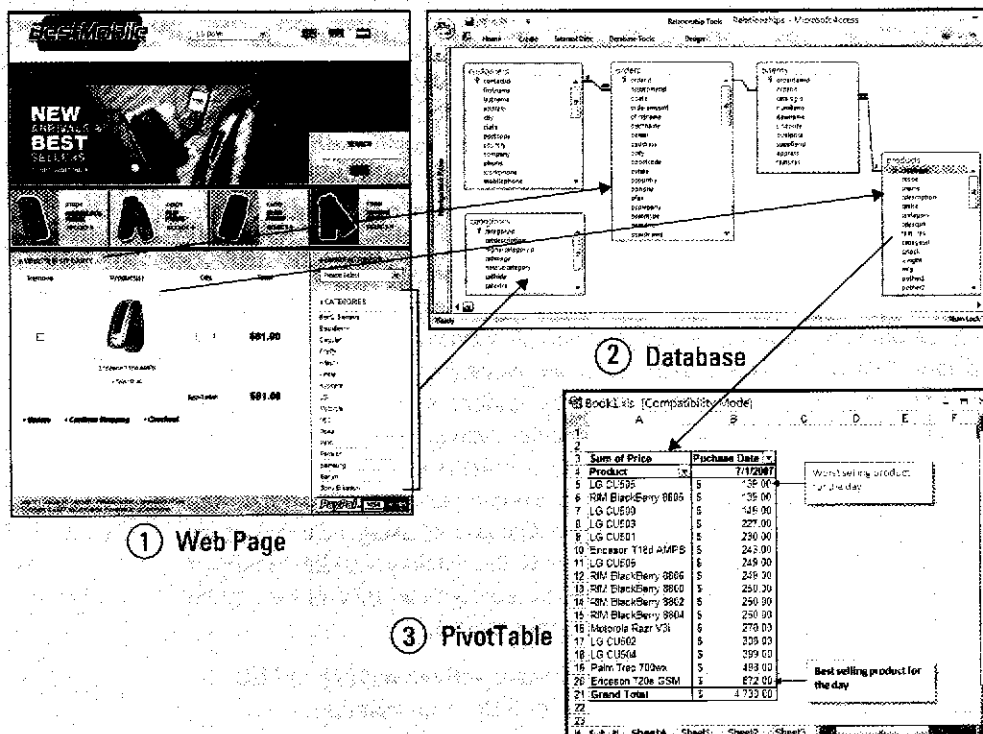


FIGURE 6.13

Zappos.com—A Data-Driven Website

FIGURE 6.14

BI in a Data-Driven Website



- **Easy to store large amounts of data:** Data-driven websites can keep large volumes of information organized. Website owners can use templates to implement changes for layouts, navigation, or website structure. This improves website reliability, scalability, and performance.
- **Easy to eliminate human errors:** Data-driven websites trap data-entry errors, eliminating inconsistencies while ensuring that all information is entered correctly.

Zappos credits its success as an online shoe retailer to its vast inventory of nearly 3 million products available through its dynamic data-driven website. The company built its data-driven website catering to a specific niche market: consumers who were tired of finding that their most-desired items were always out of stock at traditional retailers. Zappos' highly flexible, scalable, and secure database helped it rank as the most available Internet retailer. Figure 6.13 displays the Zappos data-driven website illustrating a user querying the database and receiving information that satisfies the user's request.⁶

Companies can gain valuable business knowledge by viewing the data accessed and analyzed from their websites. Figure 6.14 displays how running queries or using analytical tools, such as a PivotTable, on the database that is attached to the website can offer insight into the business, such as items browsed, frequent requests, items bought together, and so on.

section 6.2 | Business Intelligence

LEARNING OUTCOMES

- 6.5** Identify the advantages of using business intelligence to support managerial decision making.
- 6.6** Define data warehousing and data marts and explain how they support business decisions.
- 6.7** Describe the three organizational methods for analyzing big data.

SUPPORTING DECISIONS WITH BUSINESS INTELLIGENCE

Many organizations today find it next to impossible to understand their own strengths and weaknesses, let alone their biggest competitors, because the enormous volume of organizational data is inaccessible to all but the MIS department. Organization data include far more than simple structured data elements in a database; the set of data also includes unstructured data such as voice mail, customer phone calls, text messages, video clips, and numerous new forms of data such as tweets from Twitter.

The Problem: Data Rich, Information Poor

An ideal business scenario would be as follows. As a business manager on his way to meet with a client reviews historical customer data, he realizes that the client's ordering volume has substantially decreased. As he drills down into the data, he notices the client had a support issue with a particular product. He quickly calls the support team to find out all of the information and learns that a replacement for the defective part can be shipped in 24 hours. In addition, he learns that the client has visited the website and requested information on a new product line. Armed with all this information, the business manager is prepared for a productive meeting with his client. He now understands the client's needs and issues, and he can address new sales opportunities with confidence.

For many companies, the preceding example is simply a pipe dream. Attempting to gather all of the client information would actually take hours or even days to compile. With so much data available, it is surprisingly hard for managers to get information, such as inventory levels, past order history, or shipping details. Managers send their information requests to the MIS department where a dedicated person compiles the various reports. In some situations, responses can take days, by which time the information may be outdated and opportunities lost. Many organizations find themselves in the position of being data rich and information poor.

LO 6.5: Identify the advantages of using business intelligence to support managerial decision making.

BUSINESS DRIVEN INNOVATION

News Dots

Gone are the days of staring at boring spreadsheets and trying to understand how the data correlate. With innovative data visualization tools, managers can arrange different ways to view the data, providing new forms of pattern recognition not offered by simply looking at numbers. *Slate*, a news publication, developed a new data visualization tool called News Dots, that offers readers a different way of viewing the daily news through trends and patterns. The News Dots tool scans about 500 stories a day from major publications and then tags the content with important keywords such as people, places, companies, and topics. Surprisingly, the majority of daily news overlaps as the people, places, and stories are frequently connected. Using News Dots, you can visualize how the news fits together, almost similar to a giant social network. News Dots uses circles (or dots) to represent the tagged content and arranges them according to size. The more frequently a certain topic is tagged, the larger the dot and its relationship to other dots. The tool is interactive and users simply click a dot to view which stories mention that topic and which other topics it connects to in the network such as a correlation among the U.S. government, Federal Reserve, Senate, bank, and Barack Obama.⁷

How can data visualization help identify trends? What types of business intelligence could you identify if your college used a data visualization tool to analyze student information? What types of business intelligence could you identify if you used a data visualization tool to analyze the industry in which you plan to compete?

Even in today's electronic world, managers struggle with the challenge of turning their business data into business intelligence.

The Solution: Business Intelligence

Employee decisions are numerous, and they include providing service information, offering new products, and supporting frustrated customers. Employees can base their decisions on data, experience, or knowledge and preferably a combination of all three. Business intelligence can provide managers with the ability to make better decisions. A few examples of how different industries use business intelligence include:

- ✦ **Airlines:** Analyze popular vacation locations with current flight listings.
- ✦ **Banking:** Understand customer credit card usage and nonpayment rates.
- ✦ **Health care:** Compare the demographics of patients with critical illnesses.
- ✦ **Insurance:** Predict claim amounts and medical coverage costs.
- ✦ **Law enforcement:** Track crime patterns, locations, and criminal behavior.
- ✦ **Marketing:** Analyze customer demographics.
- ✦ **Retail:** Predict sales, inventory levels, and distribution.
- ✦ **Technology:** Predict hardware failures.

Figure 6.15 displays how organizations using BI can find the cause to many issues and problems simply by asking "Why?" The process starts by analyzing a report such as sales amounts by quarter. Managers will drill down into the report looking for why sales are up or why sales are down. Once they understand why a certain location or product is experiencing an increase in sales, they can share the information in an effort to raise enterprisewide sales. Once they understand the cause for a decrease in sales, they can take effective action

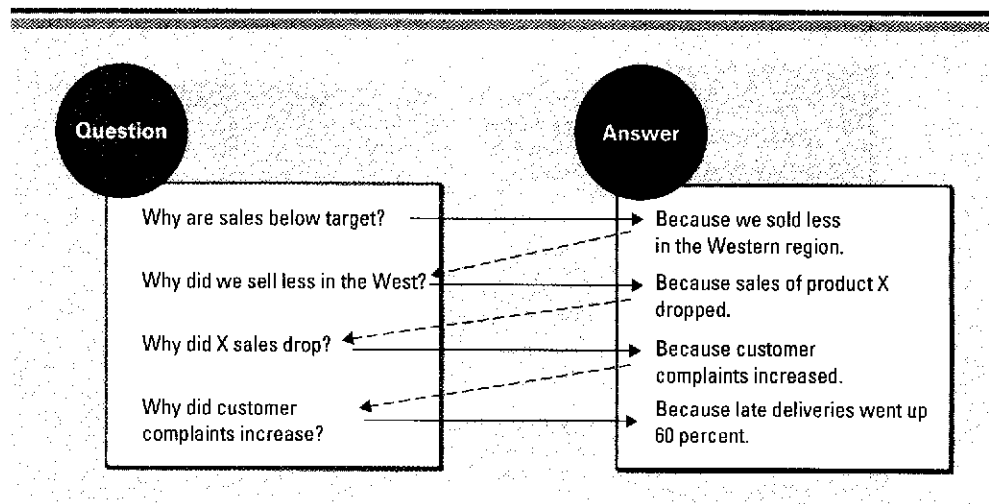


FIGURE 6.15

How BI Can Answer Tough Customer Questions

to resolve the issue. Here are a few examples of how managers can use BI to answer tough business questions:

- ✱ **Where has the business been?** Historical perspective offers important variables for determining trends and patterns.
- ✱ **Where is the business now?** Looking at the current business situation allows managers to take effective action to solve issues before they grow out of control.
- ✱ **Where is the business going?** Setting strategic direction is critical for planning and creating solid business strategies.

Ask a simple question—such as who is my best customer or what is my worst-selling product—and you might get as many answers as you have employees. Databases, data warehouses, and data marts can provide a single source of “trusted” data that can answer questions about customers, products, suppliers, production, finances, fraud, and even employees. They can also alert managers to inconsistencies or help determine the causes and effects of enterprise-wide business decisions. All business aspects can benefit from the added insights provided by business intelligence, and you, as a business student, will benefit from understanding how MIS can help you make intelligent decisions.

THE BUSINESS BENEFITS OF DATA WAREHOUSING

In the 1990s as organizations began to need more timely information about their business, they found that traditional management information systems were too cumbersome to provide relevant information efficiently and effectively. Most of the systems were in the form of operational databases that were designed for specific business functions, such as accounting, order entry, customer service, and sales, and were not appropriate for business analysis for the reasons shown in Figure 6.16.

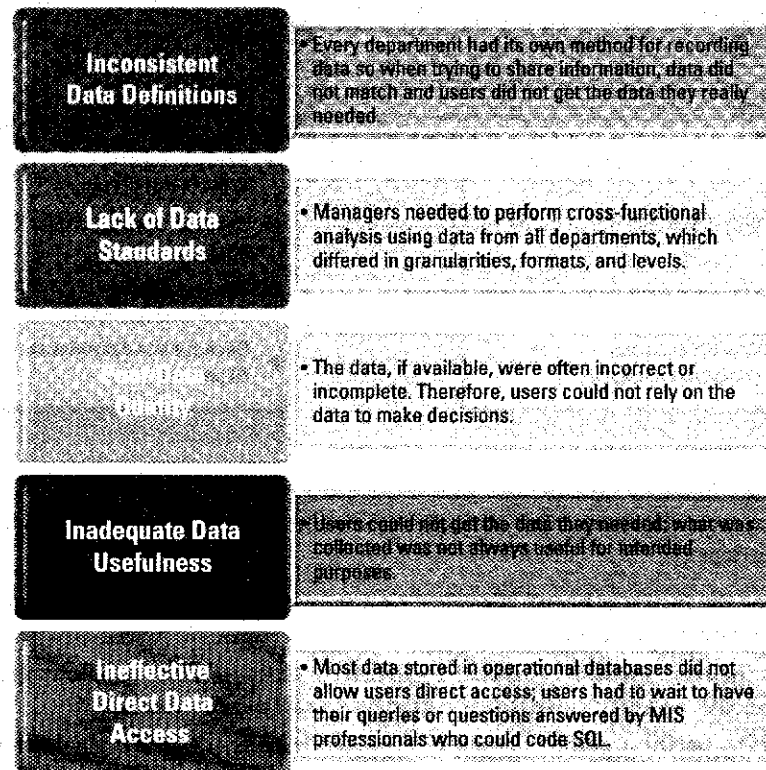
During the latter half of the 20th century, the numbers and types of operational databases increased. Many large businesses found themselves with information scattered across multiple systems with different file types (such as spreadsheets, databases, and even word processing files), making it almost impossible for anyone to use the information from multiple sources. Completing reporting requests across operational systems could take days or weeks using antiquated reporting tools that were ineffective for running a business. From this idea, the data warehouse was born as a place where relevant information could be stored and accessed for making strategic queries and reports.

A **data warehouse** is a logical collection of information, gathered from many operational databases, that supports business analysis activities and decision-making tasks. The primary purpose of a data warehouse is to combine information, more specifically, strategic information, throughout an organization into a single repository in such a way that the people who need that information can make decisions and undertake business analysis. A key idea within

LO 6.6: Define data warehousing and data marts and explain how they support business decisions.

FIGURE 6.16

Reasons Business Analysis
Is Difficult from Operational
Databases



data warehousing is to collect information from multiple systems in a common location that uses a universal querying tool. This allows operational databases to run where they are most efficient for the business, while providing a common location using a familiar format for the strategic or enterprisewide reporting information.

Data warehouses go even a step further by standardizing information. Gender, for instance can be referred to in many ways (Male, Female, M/F, 1/0), but it should be standardized on a data warehouse with one common way of referring to each data element that stores gender (M/F). Standardization of data elements allows for greater accuracy, completeness, and consistency and increases the quality of the information in making strategic business decisions. The data warehouse then is simply a tool that enables business users, typically managers, to be more effective in many ways, including:

- ✧ Developing customer profiles.
- ✧ Identifying new-product opportunities.
- ✧ Improving business operations.
- ✧ Identifying financial issues.
- ✧ Analyzing trends.
- ✧ Understanding competitors.
- ✧ Understanding product performance. (See Figure 6.17 for the three core concepts of data warehousing.)

Businesses collect a tremendous amount of transactional information as part of their routine operations. Marketing, sales, and other departments would like to analyze these data to understand their operations better. Although databases store the details of all transactions (for instance, the sale of a product) and events (hiring a new employee), data warehouses store that same information but in an aggregated form more suited to supporting decision-making tasks. Aggregation, in this instance, can include totals, counts, averages, and the like.

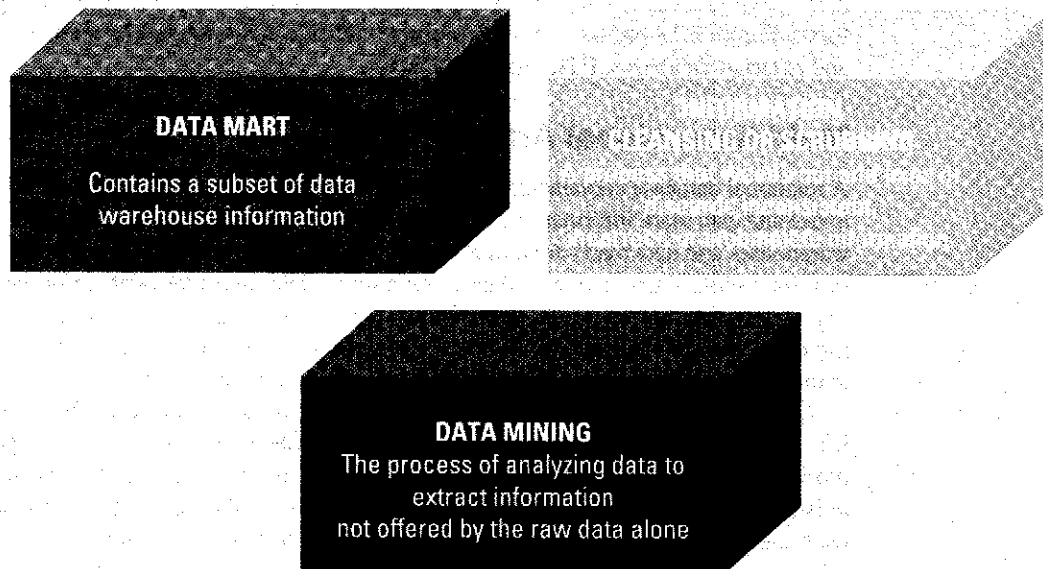


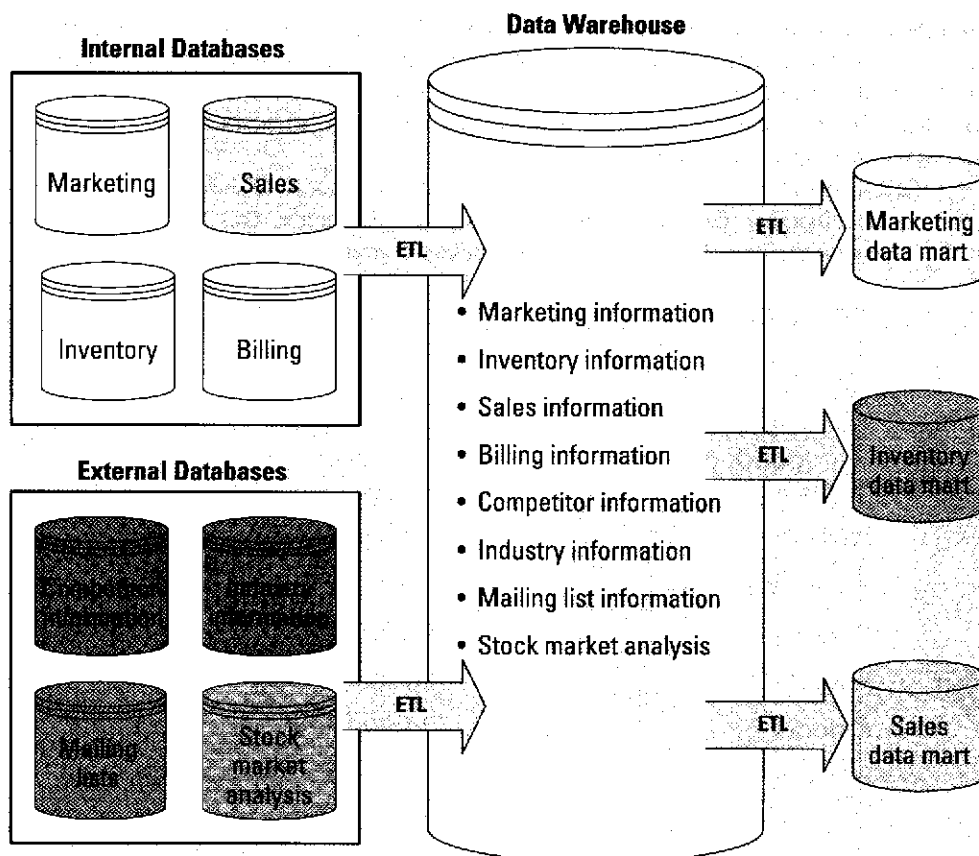
FIGURE 6.17

Three Core Concepts of Data Warehousing

The data warehouse modeled in Figure 6.18 compiles information from internal databases (or transactional and operational databases) and external databases through extraction, transformation, and loading. *Extraction, transformation, and loading (ETL)* is a process that extracts information from internal and external databases, transforms it using a common set of enterprise definitions, and loads it into a data warehouse. The data warehouse then sends portions (or subsets) of the information to data marts. A *data mart* contains a subset of data

FIGURE 6.18

Data Warehouse Model



warehouse information. To distinguish between data warehouses and data marts, think of data warehouses as having a more organizational focus and data marts as having a functional focus. Figure 6.18 provides an illustration of a data warehouse and its relationship to internal and external databases, ETL, and data marts.

Multidimensional Analysis

A relational database contains information in a series of two-dimensional tables. In a data warehouse and data mart, information contains layers of columns and rows. For this reason, most data warehouses and data marts are multidimensional databases. A dimension is a particular attribute of information. Each layer in a data warehouse or data mart represents information according to an additional dimension. An *information cube* is the common term for the representation of multidimensional information. Figure 6.19 displays a cube (cube a) that represents store information (the layers), product information (the rows), and promotion information (the columns).

After creating a cube of information, users can begin to slice and dice the cube to drill down into the information. The second cube (cube b) in Figure 6.19 displays a slice representing promotion II information for all products at all stores. The third cube (cube c) in Figure 6.19 displays only information for promotion III, product B, at store 2. By using multidimensional analysis, users can analyze information in a number of ways and with any number of dimensions. Users might want to add dimensions of information to a current analysis, including product category, region, and even forecasted versus actual weather. The true value of a data warehouse is its ability to provide multidimensional analysis that allows users to gain insights into their information.

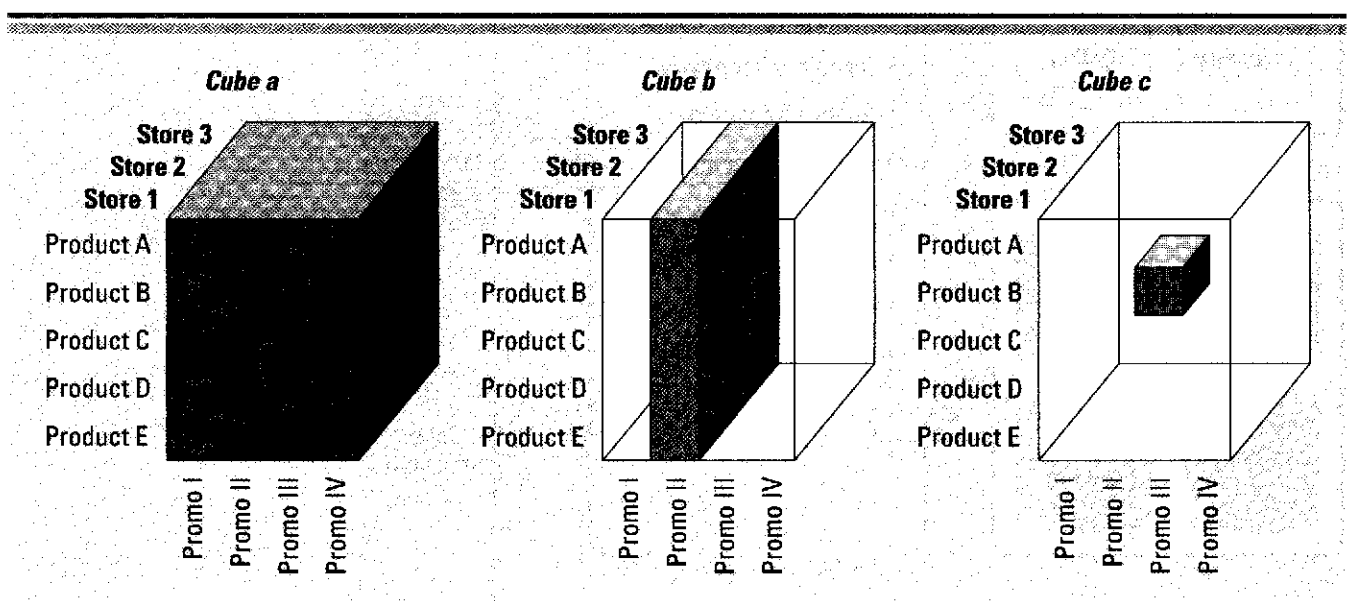
Data warehouses and data marts are ideal for off-loading some of the querying against a database. For example, querying a database to obtain an average of sales for Product B at Store 2 while Promotion III is under way might create a considerable processing burden for a database, increasing the time it takes another person to enter a new sale into the same database. If an organization performs numerous queries against a database (or multiple databases), aggregating that information into a data warehouse will be beneficial.

Information Cleansing or Scrubbing

Dirty data is erroneous or flawed data (see Figure 6.20). The complete removal of dirty data from a source is impractical or virtually impossible. According to Gartner Inc., dirty data is a business problem, not an MIS problem. Over the next two years, more than 25 percent of critical data in *Fortune* 1000 companies will continue to be flawed; that is, the information will be inaccurate, incomplete, or duplicated.

Obviously, maintaining quality information in a data warehouse or data mart is extremely important. To increase the quality of organizational information and thus the effectiveness of decision making, businesses must formulate a strategy to keep information clean.

FIGURE 6.19
A Cube of Information for
Performing a Multidimensional
Analysis on Three Stores
for Five Products and Four
Promotions



BUSINESS DRIVEN DISCUSSION

The butterfly effect, an idea from chaos theory in mathematics, refers to the way a minor event—like the movement of a butterfly's wing—can have a major impact on a complex system like the weather. Dirty data can have the same impact on a business as the butterfly effect. Organizations depend on the movement and sharing of data throughout the organization, so the impact of data quality errors are costly and far-reaching. Such data issues often begin with a tiny mistake in one part of the organization, but the butterfly effect can produce disastrous results, making its way through MIS systems to the data warehouse and other enterprise systems. When dirty data or low-quality data enters organizational systems, a tiny error such as a spelling mistake can lead to revenue loss, process inefficiency, and failure to comply with industry and government regulations. Explain how the following errors can affect an organization:

- A cascading spelling mistake
- Inaccurate customer records
- Incomplete purchasing history
- Inaccurate mailing address
- Duplicate customer numbers for different customers

Butterfly Effects

Information cleansing or scrubbing is a process that weeds out and fixes or discards inconsistent, incorrect, or incomplete information.

Specialized software tools exist that use sophisticated procedures to analyze, standardize, correct, match, and consolidate data warehouse information. This step is vitally important because data warehouses often contain information from several databases, some of which can be external to the organization. In a data warehouse, information cleansing occurs first during the ETL process and again once the information is in the data warehouse. Companies can

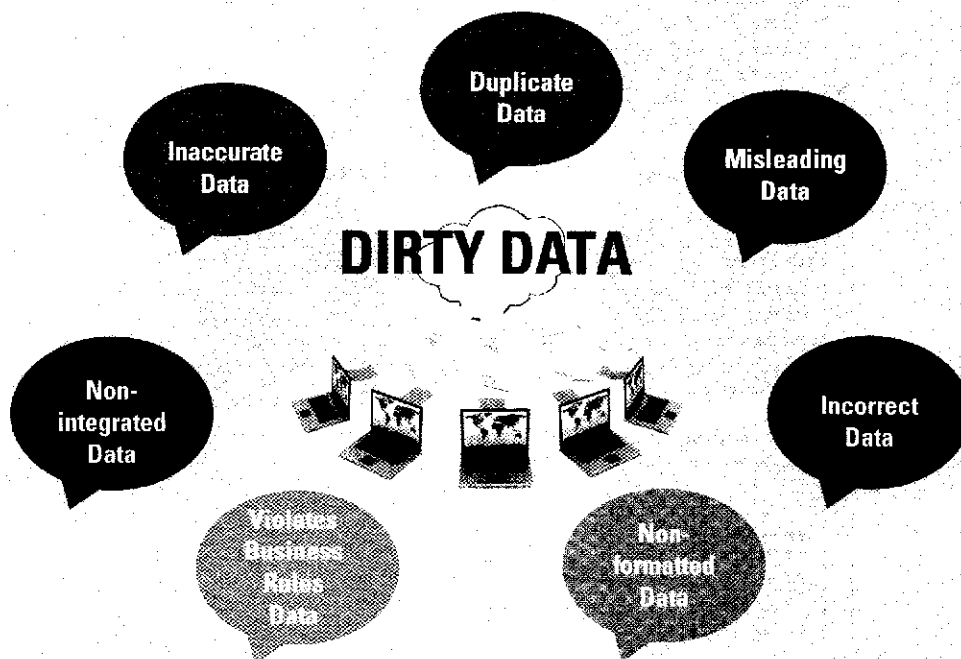


FIGURE 6.20

Dirty Data Problems

choose information cleansing software from several vendors, including Oracle, SAS, Ascential Software, and Group 1 Software. Ideally, scrubbed information is accurate and consistent.

Looking at customer information highlights why information cleansing is necessary. Customer information exists in several operational systems. In each system, all the details could change—from the customer ID to contact information—depending on the business process the user is performing (see Figure 6.21).

Figure 6.22 displays a customer name entered differently in multiple operational systems. Information cleansing allows an organization to fix these types of inconsistencies in the data warehouse. Figure 6.23 displays the typical events that occur during information cleansing.

FIGURE 6.21

Contact Information in Operational Systems

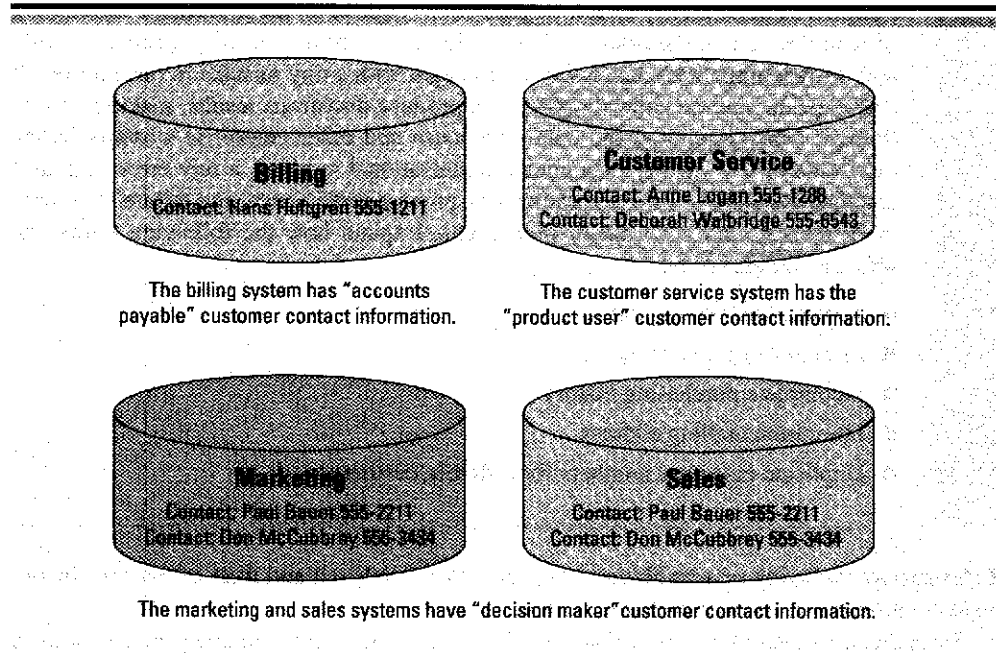
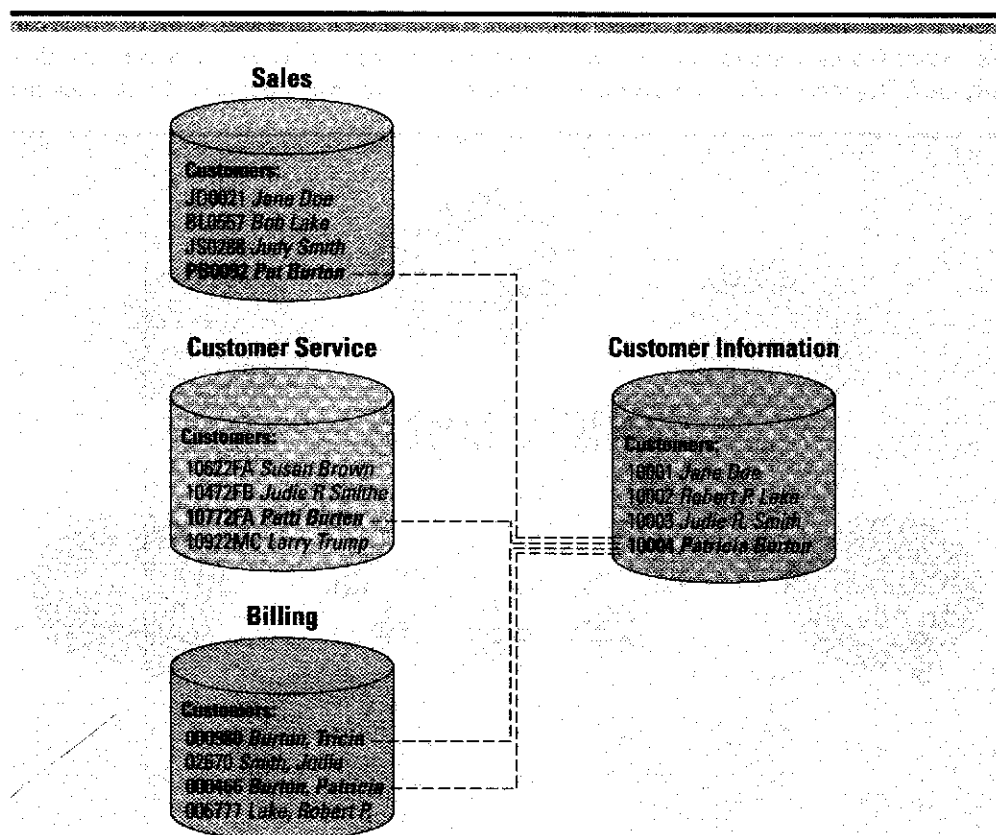


FIGURE 6.22

Standardizing a Customer Name in Operational Systems



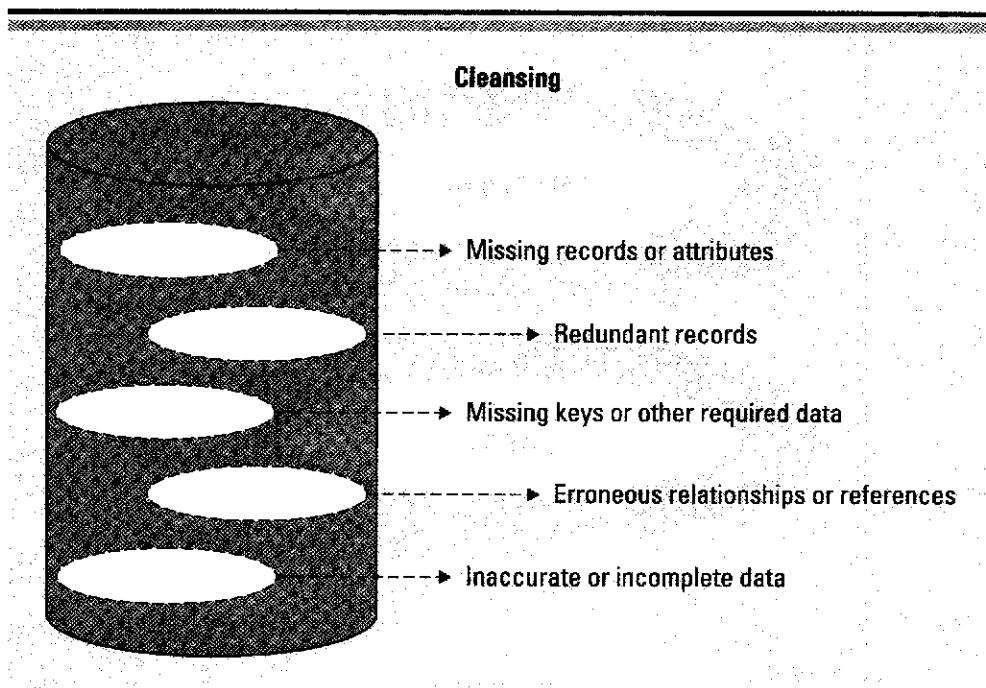


FIGURE 6.23
Information Cleansing Activities

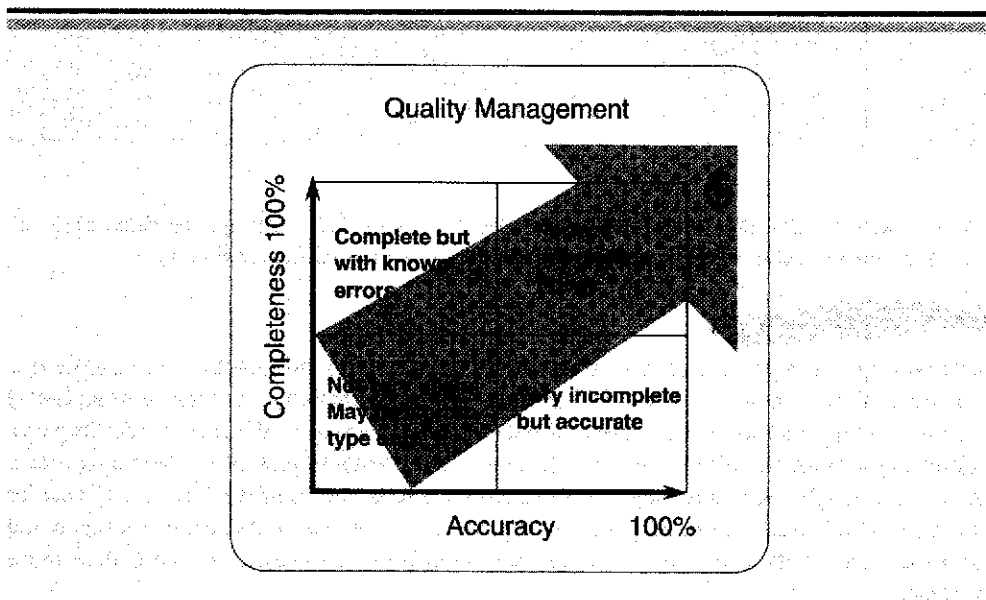


FIGURE 6.24
The Cost of Accurate and Complete Information

Achieving perfect information is almost impossible. The more complete and accurate a company wants its information to be, the more it costs (see Figure 6.24). Companies may also trade accuracy for completeness. Accurate information is correct, whereas complete information has no blanks. A birth date of 2/31/10 is an example of complete but inaccurate information (February 31 does not exist). An address containing Denver, Colorado, without a zip code is an example of accurate information that is incomplete. Many firms complete *data quality audits* to determine the accuracy and completeness of its data. Most organizations determine a percentage of accuracy and completeness high enough to make good decisions at a reasonable cost, such as 85 percent accurate and 65 percent complete.

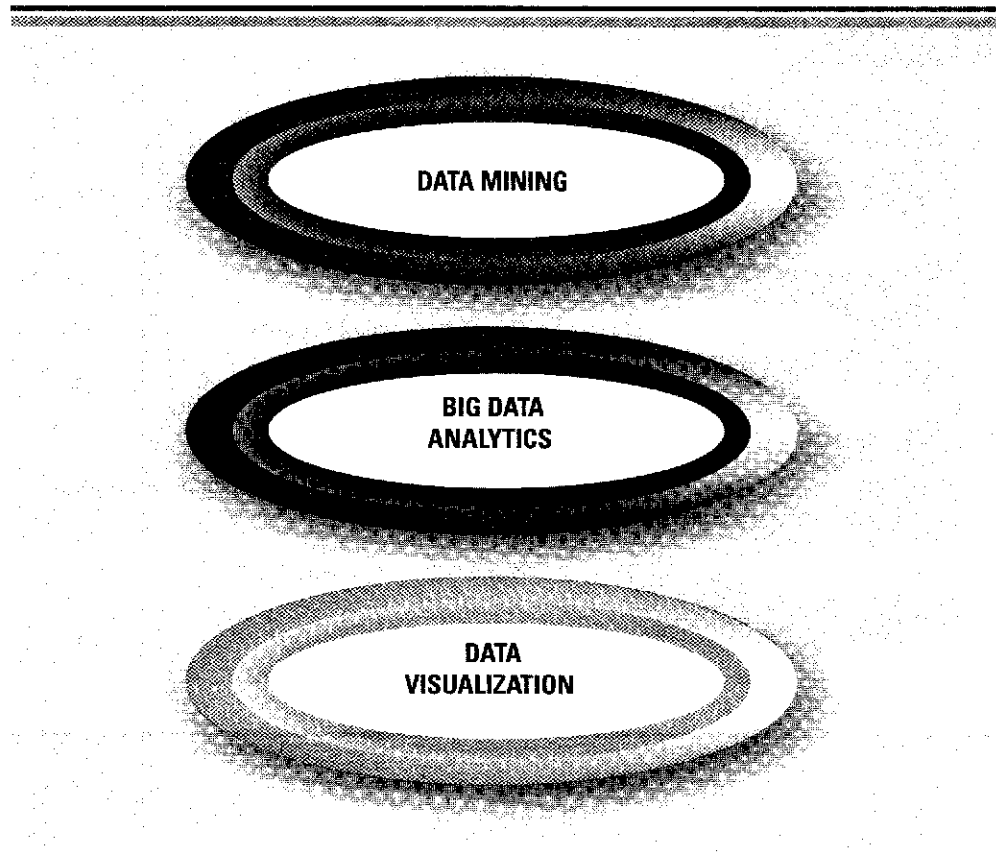
THE POWER OF BIG DATA ANALYTICS

Companies are collecting more data than ever. Historically, data were housed in functional systems that were not integrated, such as customer service, finance, and human resources. Today companies can gather all of the functional data together by the zettabyte, but finding

LO 6.7: Describe the three organizational methods for analyzing big data.

FIGURE 6.25

Three Organizational Methods
for Analyzing Big Data



a way to analyze the data is incredibly challenging. Figure 6.25 displays the three methods organizations are using to dissect, analyze, and understand organizational data.

Data mining is the process of analyzing data to extract information not offered by the raw data alone. Data mining can also begin at a summary information level (coarse granularity) and progress through increasing levels of detail (drilling down) or the reverse (drilling up). Companies use data-mining techniques to compile a complete picture of their operations, all within a single view, allowing them to identify trends and improve forecasts. Consider Best Buy, which used data-mining tools to identify that 7 percent of its customers accounted for 43 percent of its sales, so the company reorganized its stores to accommodate those customers.

To perform data mining, users need data-mining tools. **Data-mining tools** use a variety of techniques to find patterns and relationships in large volumes of information that predict future behavior and guide decision making. Data mining uncovers trends and patterns, which analysts use to build models that, when exposed to new information sets, perform a variety of information analysis functions. Data-mining tools for data warehouses help users uncover business intelligence in their data. Figure 6.26 displays the data-mining analysis methods used to uncover patterns and trends for business analysis such as:

- ✦ Analyzing customer buying patterns to predict future marketing and promotion campaigns.
- ✦ Building budgets and other financial information.
- ✦ Detecting fraud by identifying deceptive spending patterns.
- ✦ Finding the best customers who spend the most money.
- ✦ Keeping customers from leaving or migrating to competitors.
- ✦ Promoting and hiring employees to ensure success for both the company and the individual.