Chapter Outline
10.1 Data Management: A Critical Success Factor
10.2 Data (Information) Warehousing, Mining, and Analysis
10.3 Data Visualization Technologies
10.4 Marketing Databases in Action
10.5 Knowledge Management

Learning Objectives
After studying this chapter you will be able to:
1. Recognize the importance of data, their management issues, and their life cycle.
2. Describe the sources of data, their collection, and quality issues.
3. Relate data management to multimedia and document management.
4. Explain the operation of data warehousing and its role in decision support.
5. Understand the data access and analysis problem and the data mining and online analytical processing solutions.
6. Describe data presentation methods and compare and contrast geographical information systems with traditional reports.
7. Discuss the role and provide examples of database marketing.
8. Explain the concept of knowledge management.
SEARS ROEBUCK AND Company, the largest department store chain and the third largest retailer in the United States, was caught by surprise in the 1980s as shoppers defected to specialty stores and discount mass merchandisers, causing the firm to lose market share rapidly. In an attempt to change the situation, Sears used several response strategies, ranging from introducing its own specialty stores (such as Sears Hardware) to reengineering its mall-based stores. Accomplishing this goal required the retooling of IT.

Sears had 18 data centers, one in each of 10 geographical regions as well as one each for marketing, finance, and other departments. The first problem was created when the reorganization effort produced only seven geographical regions. Frequent mismatches between accounting and sales figures and information scattered among numerous databases forced users to query multiple systems even when an answer to a simple query was required. Furthermore, users found that data that were already summarized made it difficult to conduct analysis in the desired fashion and detail. Finally, errors were virtually inevitable when calculations were based on data from several sources.

To solve these problems Sears constructed a single sales information data warehouse. This replaced the 18 old databases which were packed with redundant, conflicting, and sometimes obsolete data. The new data warehouse is a simple depository of relevant decision-making data such as authoritative data for key performance indicators, sales inventories, and profit margins. Sears, known for embracing IT on a dramatic scale, completed the data warehouse and its IT reengineering efforts in under 1 year—a perfect IT turnaround story.

Using an NCR enterprise server, the 1.7 terabyte (1 terabyte equals 1 trillion bytes) data warehouse is a part of a project dubbed the Strategic Performance Reporting System (SPRS). SPRS includes comprehensive sales data; information on inventory in stores, in transit, and at distribution centers; and cost per item. This has enabled Sears to track sales by individual items in each of its 1950 stores (including 810 mall-based stores) in the United States and 1600 international stores and catalog outlets. Thus, daily margin by item per store can be easily computed, for example. Furthermore, Sears now fine tunes its buying, merchandising, and marketing strategies with previously unattainable precision.

SPRS is open to all authorized employees, who now can view each day’s sales from a multidimensional perspective (by region, district, store, product line, and individual item). Users can specify any starting and ending dates for special sales reports, and all data can be accessed via a highly user-friendly graphical interface. Sears staffers can now monitor the precise impact of advertising, weather, and other factors on sales of specific items. This means that buyers and other specialists can adjust inventory quantities, merchandising, and frequency and placement, along with a myriad of other variables, almost immediately, so they can respond quickly to environmental changes. SPRS users can also group together widely divergent kinds of products, for example, tracking sales of items marked as “gifts under $25.” Advertising staffers can follow so-called “great items,” drawn from vastly different departments, that are splashed on the covers of promotional circulars.

The ability to monitor sales by item per store enables Sears to create a sharp local market focus. For example, Sears keeps different shades of paint colors in different cities to meet local demands. Because assortments are customized in different locations the retailer comes out a winner.

By the end of 1997, the data warehouse had been used daily by over 3000 buyers, replenishers, marketers, strategic planners, logistics and finance staffers, and store managers. Response time to queries has dropped from days to minutes for typical requests. Overall, the strategic impact of the data warehouse is that it offers everybody a tool for making better decisions, and Sears retailing profits have climbed more than 20 percent annually since SPRS was implemented.

10.1 DATA MANAGEMENT: A CRITICAL SUCCESS FACTOR

THE DIFFICULTIES AND THE PROCESS

The opening case illustrates the importance of managing data in the revitalization of a large retailer. Sears consolidated these data in one place, the data warehouse, which, according to Business Week (July 31, 1995, p. 61), is the “biggest trend in information management.”

**THE DIFFICULTIES.** Data warehousing is only one facet of managing data in organizations, which is a difficult task for the following reasons:

- The amount of data increases exponentially; much past data must be kept for a long time, and new data are added rapidly.
- Data are scattered throughout organizations and are collected by many individuals using several methods and devices.
- Only small portions of an organization’s data are relevant for specific decisions.
- An ever-increasing amount of external data needs to be considered in making organizational decisions.
- Raw data may be stored in different computing systems, formats, and human and computer languages.
- Legal requirements relating to data differ among countries and change frequently.
- Selecting data management tools can be a major problem because of the huge number of products available.
- Data security, quality, and integrity are critical yet are easily jeopardized.

These difficulties, and the critical need for timely and accurate information, have prompted organizations to search for effective and efficient data management solutions. Historically, data management has been geared to supporting transaction processing by organizing the data in a hierarchical format. This format supports efficient high-volume processing; however, it is inefficient for queries and other ad hoc applications. Therefore, relational databases were added to support end-user computing and decision support. With the introduction of client–server environments, databases became distributed throughout organizations, creating problems in finding data quickly and easily. This was the major reason that Sears sought the creation of the data warehouse. As we will see later, the Intranet and extranets can also be used to improve data management.

It is now well recognized that data are a burden, but their use, in terms of information and knowledge, is power. The purpose of appropriate data management is to ease the burden and to enhance the power. The data warehouse is only one building block of effective data management. In this chapter we will look at several others. For example, Sears created a marketing database that supported enterprise-level marketing analysis applications. Recently, intranets have been playing a greater role in the support of information sharing across the enterprise, and Internet’s databases are used by almost every organization.

**DATA LIFE CYCLE PROCESS AND KNOWLEDGE DISCOVERY.** To better understand how to manage data and knowledge in organizations, it is necessary to trace how and where data flows in them. Businesses do not run on data. They run on information and their knowledge of how to put that information to use successfully.

Knowledge fuels results. Everything from innovative product designs to brilliant competitive moves relies on knowledge. Therefore, knowledge has
always been an underlying component of business. Knowledge is not readily available, however, especially in today's rapidly changing world. In many cases knowledge is continuously derived from data. However, because of the difficulties cited earlier, such a derivation may not be simple or easy.

The transformation of data into knowledge may be accomplished in several ways. In general, it is a process that starts with data collection from various sources (see Figure 10.1). These data are stored in a database(s). Then the data can be preprocessed and stored in a data warehouse. To discover knowledge, the processed data may go through a transformation that makes them ready for analysis. The analysis is done with data mining tools, which look for patterns, and intelligent systems, which support data interpretation. The result of all these activities is generated knowledge. Both the data, at various times during the process, and the knowledge, derived at the end of the process, may need to be presented to users. Such a presentation can be accomplished several times, using different presentation tools. The created knowledge may be stored in a knowledge base. This process can guide us through the content of the remaining sections of this chapter.

**DATA SOURCES, COLLECTION, AND QUALITY**

The data life cycle begins with the acquisition of data from data sources. Data can include documents, pictures, maps, sound, and animation and they can be stored and organized in different ways before and after use. Data also include concepts, thoughts, and opinions and can be raw or summarized. Many IS applications use summary or extracted data. Data can be classified as internal, personal, and external.

**INTERNAL DATA.** An organization's internal data are stored in one or more places. These data are about people, products, services, and processes. For example, data about employees and their pay are usually stored in the corporate database. Data about equipment and machinery may be stored in the maintenance department database. Sales data can be stored in several places, aggregate sales data in the corporate database and details at each regional database, as in the Sears case. Internal data are usually accessible via an organization's computer network system.

**PERSONAL DATA.** IS users or other corporate employees may document their own expertise by creating personal data. These include, for example, subjective estimates of sales, opinions about what competitors are likely to do, and certain rules and formulas developed by the users. These data can reside on the user's PC.
or be placed on some public database (some personal data are not documented but are kept in people's memories).

**EXTERNAL DATA.** There are many sources for external data. They range from commercial databases to sensors and satellites. Data are available on CD-ROMs, on the Internet servers, as films, and as music or voices. Pictures, diagrams, atlases, and television are also sources of data. Government reports constitute a major source for external data (see Box 10.1) Hundreds of thousands of organizations worldwide place publicly accessible data on their Web servers, flooding us with data. Most external data are irrelevant to a specific application. Yet much external data must be monitored and captured to ensure that important data are not overlooked. Large amounts of external data are available on the Internet.

---

**A Closer Look**  
**BOX 10.1**  
**Sources of Business External Data—A Sampler; Most Available on the Internet**

Thousands of public databases exist, most of which are available on the Internet.

**FEDERAL PUBLICATION**
Survey of Current Business (Department of Commerce) (continues Business Conditions Digest in short form)—Monthly, general business conditions  
Monthly Labor Review (Department of Labor)—Monthly employment statistics (a journal with articles)  
Federal Reserve Bulletin (FEDRB)—Business and economics papers and statistics; government regulations  
Employment and Earnings (Department of Labor)—Monthly; more detailed than Monthly Labor Review  
Commerce Business Daily (Department of Commerce)—Daily announcements and news provided by the U.S. Department of Commerce

**OTHER**
International Monetary Fund—Report of balance of payments, including currency rates, for participating countries  
Moody’s—A series of manuals including abstracted information and balance sheets of most large U.S. corporations; intended for investors  
Thomas Register Online—Information on over 150,000 U.S. companies and their products  
Standard & Poor’s—Periodically updated report of financial stability of most U.S. corporations  
Advertising Age—Marketing newspaper, with a great deal of data on marketing  
Annual Editor & Publisher Market Guide—Annual report of marketing information by standard metropolitan statistical area (SMSA)  
DIALOG and LEXIS-NEXIS—The major players in the field, providing access to several hundred databases in many disciplines  
ABI/Inform—Available on CD-ROM in many libraries; covers hundreds of business journals for abstracts and full text

**INDEXES** (now available for a fee on the internet)
Encyclopedia of Business Information Sources—Gale Research, Inc.; updated annually; bibliographic guide on about 1,000 business subjects, including online databases
Encyclopedia of Information Systems and Services—Gale Research, Inc.; updated annually; descriptive guide to databases in electronic form
The CD-ROM Directory—TFPL Publishing; updated annually; index of CD-ROM databases

---

**The Internet and Commercial Database Services**

Some external data flow to an organization on a regular basis through the electronic data interchange (EDI) or through other traditional company-to-company channels, but much data are accessible via the Internet.

- **The Internet.** Many thousands of databases all over the world are accessible through the Internet. A user can access home pages of vendors, clients, and
competitors. He or she can view and download information or conduct research. The Internet, as discussed in Chapter 5, is becoming the major source of external data for many decision situations (see Vassos, 1996, for a comprehensive list of Internet data sources).

**Commercial online publishing.** An *online publisher* sells access to specialized databases, newspapers, magazines, bibliographies, and reports. Such a service can provide external data to users in a timely manner and at a reasonable cost. Several thousand services are currently available, most of which are accessible via the Internet. Representative examples of providers can be found in A Closer Look 10.1; information related to investors is provided in Chapter 6.

**METHODS FOR COLLECTING RAW DATA**

The diversity of data and the multiplicity of sources make the task of data collection fairly complex, creating quality and integrity problems. Sometimes it is necessary to collect raw data in the field. In other cases it is necessary to elicit data from people. Raw data can be collected manually or by instruments and sensors. Data can also be scanned or transferred electronically. Representative manual data collection methods are time studies, surveys, observations, and contributions from experts.

Although a wide variety of hardware and software exists for data storage, communication, and presentation, much less effort has gone into developing software tools for data capture in environments where complex and unstable data exist. Insufficient methods for dealing with such situations may serve to limit the effectiveness of IS development and use.

The collection of data from multiple external sources may be an even more complicated task. One way to improve it, according to Roland (1994), is to use a data flow manager (DFM), which takes information from external sources and puts it where it is needed, when it is needed, in a usable form. A DFM consists of (1) a decision support system, (2) a central data request processor, (3) a data integrity component, (4) links to external data suppliers, and (5) the external data supplier’s processes.

Regardless of how they are collected, data need to be validated. A classic expression that sums up the situation is “garbage in, garbage out” (GIGO). Therefore, safeguards on data quality are designed to prevent data problems.

**DATA QUALITY**

**Data quality** (DQ) is an extremely important issue since quality determines the data’s usefulness as well as the quality of the decisions based on these data. Data are frequently found to be inaccurate, incomplete, or ambiguous, particularly in organizational databases. The economical and social damage from poor quality data costs of billions of dollars. An example of typical data problems, their causes, and possible solutions is provided in Table 10.1. For a discussion of data auditing and controls see Chapter 15.

Strong et al. (1997) conducted extensive research on DQ problems and categorized these problems into the following four categories and dimensions (see Wang et al., 1995, 1996):

- **Intrinsic DQ:** Accuracy, objectivity, believability, and reputation
- **Accessibility DQ:** Accessibility and access security
- **Contextual DQ:** Relevancy, value added, timeliness, completeness, amount of data
- **Representation DQ:** Interpretability, ease of understanding, concise representation, consistent representation
Strong et al. (1997) have suggested that once the major variables and relationships in each category are identified, an attempt can be made to find out how to better manage the data. For example, DQ problems in accessibility are shown in Figure 10.2. Note that some of the problems are technical ones such as capacity, while others relate to potential computer crimes, as discussed in Chapter 15.

### Table 10.1 Data Problems

<table>
<thead>
<tr>
<th>Problem</th>
<th>Typical Cause</th>
<th>Possible Solutions (in Some Cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data are not correct.</td>
<td>Raw data were entered inaccurately.</td>
<td>Develop a systematic way to ensure the accuracy of raw data. Automate (use scanners or sensors).</td>
</tr>
<tr>
<td>Data derived by an individual were generated carelessly.</td>
<td>Carefully monitor both the data values and the manner in which the data have been generated. Check for compliance with collection rules.</td>
<td></td>
</tr>
<tr>
<td>Data are not timely.</td>
<td>The method for generating the data was not rapid enough to meet the need for the data.</td>
<td>Modify the system for generating the data. Move to a client-server system. Automate.</td>
</tr>
<tr>
<td>Data are not measured or indexed properly.</td>
<td>Raw data were gathered according to a logic or periodicity that was not consistent with the purposes of the analysis.</td>
<td>Develop a system for rescaling or recombining the improperly indexed data. Use intelligent search agents.</td>
</tr>
<tr>
<td>Needed data simply do not exist.</td>
<td>No one ever stored the data needed now.</td>
<td>Whether or not it is useful now, store data for future use. Use the Internet to search for similar data. Use experts.</td>
</tr>
<tr>
<td>Required data never existed.</td>
<td></td>
<td>Make an effort to generate the data or to estimate them (use experts). Use neural computing for pattern recognition.</td>
</tr>
</tbody>
</table>


**Figure 10.2** Accessibility DQ problem pattern. From Strong et al., 1997, p. 106.
One of the major issues of DQ is **data integrity**. Older filing systems may lack integrity. That is, a change made in the file in one place may not be made in the file in another place or department. This results in conflicting data.

Data are collected on a routine basis or for a special application. In either case, it is necessary to organize and store them before they can be used, as described next.

### Multimedia and Object-Oriented Databases

The manner in which data are collected and their intended use dictate the format of their organization and storage. Data are organized and stored in files and databases. The major conventional, logical data organizations are **hierarchical, network**, and **relational**. These are described with other database topics in Appendix C. The object-oriented database is the most widely used of the newest methods of data organization.

An **object-oriented database** is a part of the object-oriented paradigm, which also includes object-oriented programming, operating systems, and modeling. Its technical details are described in Appendix C. It is used for complex applications, such as computer-integrated manufacturing, that require accessibility to pictures, blueprints, and other images, as shown in the case of Daimler-Benz.

---

### IT at Work: Object-Oriented Database at Daimler-Benz

Daimler-Benz, the German automobile and aerospace multinational corporation known for its Mercedes cars, is a pioneer in the use of object-oriented databases for business applications. The company uses this technology to integrate data stored worldwide, throughout the enterprise, and to expedite decision making. Cooperating with a small software vendor, Daimler-Benz integrated object databases with its existing, or “legacy,” information systems. The new technology enables Daimler-Benz to tie its diverse database systems together and to link to customers and suppliers in a network-based, distributed, online information system.

Object-oriented databases have been around since the mid-1980s, but they were used primarily for computer-aided design and graphical information systems. Since the early 1990s, object-oriented databases have become popular for many other business applications as a result of their ability to integrate with existing relational and hierarchical databases and with networks.

This is why Daimler-Benz pioneered the large-scale use of the technology. The company maintains an object database of product-model data based on an automotive industry standard. The databases are used in design, manufacturing, and sales of both automobile and aerospace products. One reason that Daimler-Benz was willing to invest heavily in the emerging technology was that it could meet its business goals of high productivity and quality and at the same time leverage the high investment it had in its legacy information systems.

**FOR FURTHER EXPLORATION.** Object-oriented databases were used originally for pictures and graphs. Why is Daimler-Benz interested in such capabilities, and why does it want to integrate them with the legacy systems? ▲

A considerable amount of corporate information resides outside the computer in the form of documents, maps, photos, images, and videotapes. For companies to build applications to take advantage of these rich data types, they must use a multimedia database management system. Database vendors are enhancing their offerings to include the ability to manage and manipulate multimedia data.

There are many applications of multimedia database, especially in such industries as newspapers and TV, where multimedia plays an important role, as shown in the case of Fountain New Media.

### 10.1 Data Management: A Critical Success Factor

#### Southam Interactive Database

Southam New Media, one of the largest newspaper chains in Canada, has created a multimedia database called a virtual newsstand. The Southam Interactive Database allows the entry, filing, and retrieval of data in all formats used by Southam: wire service text, newspaper and magazine text, classified ads, photos, graphics, television, video and sound, software programs, and Web content. Any unit of data has identifiable attributes that can be stored and retrieved through a single access point. Southam’s initial investment was only $500,000—much less than if its newspapers were paginated on different systems. According to Peter Irwin, president of Southam New Media, there is a single access window for retrieval and input, with one export button, with which content can be exported onto the Web or, for example, a commercial service provider such as America Online. Basically, what the development team did was to write software that allowed extraction from all Southam’s newspapers on line and tried to make sense of what was in a story. The information was then put into an Oracle database, which enabled flexibility. The system was set up on Southam’s intranet.

Southam owns Business World, a Canadian television show. The company plans to put relevant video from the program into the database using standard video-capturing technology. The video then can be attached to any story. Future applications include a full electronic distribution of Southam’s news content.

**FOR FURTHER EXPLORATION.** Why is the database called interactive, and how is the Web related to this application?


### Document Management

Electronic organization of data is interrelated with the issue of electronic document management.

There are several major problems with paper documents. For example, in document maintenance, we can pose the following questions: (1) Does everyone have the current version? (2) How often does it need to be updated? and (3) How can the distribution of documents to the appropriate individuals in a timely manner be managed? When documents are provided in electronic form from a single repository (typically a Web server), only the current version needs to be provided. Access can be restricted as required. For example, many firms maintain their telephone directories in electronic form on the Intranet to eliminate their hard-copy distribution and constant correction and redistribution. Current **document management** technology grows out of the business community, where, according to the GartnerGroup, some 85 percent of corporate information resides in documents. The need for greater efficiency in handling business documents to gain an edge on the competition has fueled the increased availability of document management systems, also known as electronic document management.

Essentially, **electronic document management** (EDM) systems provide information to decision makers in a usable format. The Thomas Cook Company,
for example, uses a document management system to handle travel refund applications. The system works on the PC desktop and has automated the workflow process, helping the firm double its volume of business while adding only about 33 percent more employees (see Cole, 1996).

Document management is the automated control of electronic documents, page images, spreadsheets, word processing documents, and complex, compound documents through their entire life cycle within an organization, from initial creation to final archiving. Document management allows organizations to exert greater control over production, storage, and distribution of documents, yielding greater efficiency in the reuse of information, the control of a document through a workflow process, and the reduction of product cycle times. The full range of functions that a document management system may perform includes document identification, storage, and retrieval; tracking; version control; workflow management; and presentation.

Document management usually includes computerized imaging systems that can result in substantial savings, as at USAA.

---

**IT at Work:**

**THE IMAGING SYSTEM AT UNITED SERVICES AUTOMOBILE ASSOCIATION**

United Services Automobile Association (USAA) is a large insurance company in San Antonio, Texas, serving about 2 million military officers, former officers, and their dependents. In 1997 the company processed about 130,000 documents every workday.

In the 1980s, the company employed 120 clerks whose only job was to search files (which occupied 39,000 square feet of office space) for information when needed. Searches for one document took anywhere from an hour to 2 weeks, and some documents were never found. However, using an environment called Automated Insurance Environment, USAA has been transformed into a completely paperless company. Since 1993, employees have been able to scan over 50 million pieces of mail per year into a computer database. Agents, using special imaging terminals, then access the database to process information and assist customers. Customers can now obtain *instant* answers to such questions as: “Is the car we bought an hour ago insured against theft?” The system is also used to expedite the treatment of claims. When a customer calls to report a car accident, the telephone call is digitized and stored. Subsequently, all documents (photos, doctors’ reports, appraisers’ reports, and so on) are scanned into the computer. Once the case is closed, it is stored on a CD-ROM for future reference.

Here is how the data entry works. Every day the company receives almost 25,000 letters as well as thousands of telephone calls; every day the company sends over 60,000 letters and policies. All the documents are indexed on an IBM mainframe and then are scanned to create a digital electronic picture. The imaged documents can then be accessed from anywhere in the company for viewing or processing. Special high-resolution terminals are used for displaying and printing the imaged documents. The company uses **electronic forms** to expedite the preparation of standard documents.

The system improves productivity of employees, reduces the cost of storing documents (most paper documents are destroyed), and improves customer service. Employee productivity is improved by eliminating the time necessary to search for the appropriate documents. With the new system, all pertinent data can be called to the user’s screen within 6 minutes of the request’s being issued. With the old system, document collection required an average of 0.5 man-hours. (If any of the documents were not housed in the local facility, the document collection averaged 1.2 hours.) At an average hourly cost of $15 for document handlers, the savings are $7.05 per document, or $70,500,000 for the 10,000,000 documents handled annually.

Most users find the imaging system easier to work with than the old paper-clogged system. The system is also used for scheduling work and monitoring workflow.

**FOR FURTHER EXPLORATION.** In what other industries can such a system be beneficial? ▲

The major tools are workflow, authoring tools, and databases (object-oriented mixed with relational, known as object-relational database management systems; see Appendix C).

In many organizations, documents are now viewed as multimedia objects with hyperlinks. The Web provides easy access to pages of information. Document management systems excel in this area. Web-enabled document management systems also make it easy to put information on an intranet, since many of them provide instantaneous conversion of documents to HTML. BellSouth, for example, saves an estimated $17.5 million each year through its intranet-enabled forms management system.

McDonnell Douglas distributes aircraft service bulletins to their customers around the world using the Internet. The company used to distribute a staggering volume of bulletins to over 200 airlines, using over 4 million pages of documentation every year. Now, it is all on the Web, saving money and time both to the company and to its customers. For details see http://home.netscape.com/comprod/at_work/customer_profile/mcdonnell.html.

One of the major vendors of document management is Lotus Development Corporation. Its document databases and their replication property provide many advantages for group work and information sharing (see Chapter 5).

10.2 DATA (INFORMATION) WAREHOUSING, MINING, AND ANALYSIS

**Transaction versus Analytical Processing**

Data processing in organizations can be viewed either as transactional or analytical. Transaction processing, which has been introduced in Chapter 2 and discussed in Chapter 8, is the routine daily processing of the transactions of the organizations such as ordering or billing. The data in such cases are organized mainly in a hierarchical structure (see Appendix C) and are processed by the IS department. The databases and the processing systems involved are known as operational systems and the results are mainly summaries and reports.

In today's fast paced and highly competitive marketplace, access to data is critical. The most successful companies are those that can respond quickly and flexibly to market changes and opportunities, and the key to this response is the effective and efficient use of data and information. It is not sufficient, therefore, to conduct transaction processing. A supplementary activity, called analytical processing, is done mainly by end users. Placing strategic information in the hands of decision makers aids productivity and empowers them to make better decisions, leading to greater competitive advantage. End users need direct access to corporate data. A good data delivery system therefore should be able to support:

- Easy data access by the end users themselves
- Quicker decision making
- More accurate and effective decision making
- Flexible decision making

Such requirements are not met by the traditional database access. Why is this so?

Analytical processing, which includes DSS, EIS, and other end-user activities, is based on data stored in the operational system and possibly on external data. There are basically two options for conducting analytical processing. One is to work directly with the operational systems (the “let’s use what we have” approach), using software tools and components known as front-end tools and middleware (see Appendix C). This option, shown in Figure 10.3, can be optimal
for companies that do not have a large number of users running queries and conducting analyses against the operating systems. It is also an option for departments that consist mainly of users who have the technical skills necessary for writing their own programs to generate ad hoc reports and conduct analysis (such as engineering). This option includes extensive use of “fourth-generation” (4GL) tools such as spreadsheets and graphics. Although it is possible to use the 4GLs as query and reporting tools, they may not be effective, flexible, or easy enough to use in many cases.

More recently, there has been a wave of front-end tools that allow end users to ease these problems by directly conducting queries and reporting on data stored in the operational databases. These tools are designed to empower end users with little or no involvement of IS personnel. The problem with this approach, however, is that the tools are only effective with end users that have a medium to high degree of knowledge about the databases—requiring nontechnical personnel, in effect, to think like database administrators. More often than not, IS personnel spend more time supporting the users of these tools than they do generating ad hoc reports and doing analytical processing by themselves. This calls for a better solution.

This solution is the second option of analytical processing and involves three concepts:

- A business representation of data for end users
- A client–server environment that gives the users query and reporting capabilities
- A server-based repository, the data warehouse, that allows centralized security and control over the data

### The Data Warehouse and Marts

The Sears case illustrates the benefits of a single depository place, called a data warehouse. The major benefits are the abilities (1) to reach data quickly, since they are located in one place, and (2) to do it easily, frequently by end users themselves. Let’s see what a data warehouse is and what its benefits are.

The purpose of a data warehouse is to establish a data repository that makes operational data accessible in a form readily acceptable for analytical processing activities such as decision support, EIS, and other end-user applications. As part of this accessibility, detail-level operational data must be transformed to a rela-
tional form, which makes them more amenable to analytical processing. Thus, data warehousing is not a concept by itself but is interrelated with data access, retrieval, analysis, and visualization.

The process of building and using a data warehouse is shown in Figure 10.4. Data are stored in operational systems (left side of the figure). Not all data are transferred to the data warehouse, and frequently only a summary of the data is transferred. The data are organized within the warehouse as a relational database so it is easy for end users to access. Also, the data are organized by subject, such as by functional area, vendor, or product. In contrast, operational data are usually organized according to a business process, such as shipping, purchasing, or inventory control. These are basically legacy application systems.

Data warehouses allow for the storage of **metadata**, which include data summaries that are easier to index and search, especially with Web tools. An optional arrangement is the creation of data marts.

A **data mart** is a subset of the data warehouse and it is dedicated to a functional or regional area. For example, many companies have a marketing data mart, or a data mart for their foreign operations. Some companies have several data marts. The data in the data mart are a replication of a portion of the data in the data warehouse (see Watterson, 1996). Data marts can be a substitution for a data warehouse or can be additional databases, holding only part of the data warehouse content. In the latter case, end users can use the warehouse and/or the marts for many applications (shown on the right side of Figure 10.4). These include query, EIS/DSS, reporting, online analytical processing (OLAP), knowledge discovery, and data mining. Some of these applications are described in Chapter 9; others are described later in this section. The end users employ a variety of 4GL tools for these applications, including spreadsheets, DBMS, and graphics.

Middleware tools enable access to the data warehouse (see Appendix C).

**CHARACTERISTICS OF DATA WAREHOUSING.** The major characteristics of data warehousing are:

1. **Organization.** Data are organized by detailed subject; e.g., by customer,
vendor, product, price level, and region, containing only information relevant for decision support.

2. **Consistency.** Data in different operational databases may be encoded differently. For example, gender data may be encoded 0 and 1 in one place and “m” and “f” in another. In the warehouse they will be coded in a consistent manner.

3. **Time variant.** The data are kept for 5 to 10 years so they can be used for trends, forecasting, and comparisons over time.

4. **Nonvolatile.** Once entered into the warehouse, data are not updated.

5. **Relational.** Typically the data warehouse uses a relational structure.

**BENEFITS.** Moving information off the mainframe presents a company with the unique opportunity to restructure its IT strategy. Companies can reinvent the way in which they shape and form their application data, empowering end users to conduct extensive analysis with them. Another immediate benefit is providing a consolidated view of corporate data, which is better than providing many smaller (and differently formatted) views. For example, separate production systems may track sales and coupon mailings. Combining data from these different systems may yield insights into the cost efficiency of coupon sales promotions that would not be immediately evident from the output data of either system alone. Integrated within a data warehouse, however, such information can be easily extracted. Another benefit is that data warehousing allows information processing to be off loaded from individual (legacy) systems onto lower cost servers. Once this is done, the end-user tools can handle a significant number of end-user information requests. Furthermore, some operational system reporting requirements can be moved to decision support systems, thus freeing up production processing.

These benefits can improve business knowledge, provide competitive advantage, enhance customer service and satisfaction, facilitate decision making, and help in streamlining business processes, as happened at the Bank of Ireland.

---

**IT at Work:**

**THE BANK OF IRELAND COMPETES USING A DATA WAREHOUSE**

Liberalized banking laws have resulted in stiff competition on a variety of products and services among banks and other financial institutions all over the world. The Bank of Ireland uses a data warehouse to exploit regulatory changes that have opened up a myriad of opportunities for revenue and profit growth. The warehouse allows managers to access corporate data directly and easily with their PCs, using Microsoft Excel and Access. This access makes for rapid turnover time in getting answers to queries and conducting analyses, such as examination of alternative responses to competitors’ activities.

Easy accessibility to data enables bank employees to analyze large amounts of information quickly, to respond rapidly to changing business conditions, and to determine strategies for gaining competitive advantage. Previously, the bank outsourced many applications. Its IS department was providing answers to queries that are now found by end users, freeing the IS personnel to do system development. According to Des Harrold, a sector manager, in a country of only 3.5 million people, companies grow by stealing competitors’ customers. In addition to increasing market share at the expense of its main competitor, the bank has saved about $800,000 a year on systems development.

**FOR FURTHER EXPLORATION.** How were banking deregulations exploited by the bank? How can the bank respond when its competitors also build warehouses? ▲

COST. The cost of a data warehouse can be very high, both to build and to maintain. Furthermore, it may be difficult and expensive to incorporate data from obsolete legacy systems. Finally, there may be a lack of incentive to share data. Therefore, a careful feasibility study must be undertaken before a commitment is made to data warehousing.

ARCHITECTURE AND PROCESS. There are several basic architectures for data warehousing. Two common ones are two-tier and three-tier architecture. In three-tier architecture, data from the warehouse are processed again and deposited in an additional multidimensional database, organized for easy multidimensional analysis and presentation (see Section 10.3), or are replicated in data marts.

PUTTING THE WAREHOUSE ON THE INTRANET. Delivery of data warehouse content to decision makers throughout the enterprise can be done via an intranet. Users can view, query, and analyze the data and produce reports using Web browsers. This is an extremely economical and effective method of delivering data (see Information Advantage, 1997).

SUITABILITY. Data warehousing is most appropriate for organizations in which:

- Large amounts of data need to be accessed by end users.
- The operational data are stored in different systems.
- An information-based approach to management is in use.
- There is a large, diverse customer base (such as in a utility company or a bank).
- The same data are represented differently in different systems.
- Data are stored in highly technical formats that are difficult to decipher.
- Extensive end-user computing is performed (many end users performing many activities).

For further discussion see McFadden and Watson (1996), Barquin and Edelstein (1997), and Inmon (1996). See also www.cait.wustl.edu/cait/papers/rpism/vol1_no1.

**Knowledge Discovery and Data Mining**

Once the data are in the data warehouse and/or data marts they can be accessed by end users. Users can then conduct several activities with the data, ranging from decision support and executive support analyses, which are discussed in Chapter 9, to **knowledge discovery**.

The process of extracting useful knowledge from volumes of data is known as **knowledge discovery in databases** (KDD) and it is the subject of extensive research (Fayyad and Uthurusamy, 1996). KDD’s objective is to identify valid, novel, potentially useful, and ultimately understandable patterns in data. The process starts with identifying which data to consider in the data warehouse and then preprocessing these data (see Figure 10.1) to be ready for analysis.

**The Foundations of KDD.** KDD techniques are the result of a long process of research and product development. This evolution began when business data were first stored on computers, continued with improvements in data access, and, more recently, generated technologies that allowed users to navigate through their data in real time. KDD is useful because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms
Commercial databases are growing at unprecedented rates. According to the Data Warehouse Institute, a 1996 survey of data warehouse projects, found that those of 19 percent of the respondents were beyond 50 gigabytes in size, while 59 percent were expected to be there within 6 months. In some industries, such as retail, these numbers can be much larger, reaching terabytes, as at Sears.

**TOOLS AND TECHNIQUES OF KDD.** Knowledge discovery has been done since the 1960s. However, the enabling techniques have been expanded and improved over time. The KDD processes have appeared under various names and have shown different characteristics. The evolution of the KDD tools over time can be divided into four major stages. These stages are shown in Table 10.2. As time passes the KDD can answer more complex business questions. The table lists the enabling technologies and their characteristics. Data access techniques are discussed in Appendix C of this book. Online analytical processing and data mining are discussed in this section; multidimensionality is discussed in Section 10.3; and massive databases, as employed in database marketing, are the subject of Section 10.4.

**Table 10.2 Steps in the Evolution of Knowledge Discovery**

<table>
<thead>
<tr>
<th>Evolutionary Step</th>
<th>Business Question</th>
<th>Enabling Technologies</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection (1960s)</td>
<td>What was my total revenue in the last five years?</td>
<td>Computers, tapes, disks</td>
<td>Retrospective, static data delivery</td>
</tr>
<tr>
<td>Data Access (1980s)</td>
<td>What were unit sales in New England last March?</td>
<td>Relational databases (RDBMS), Structured Query Language (SQL)</td>
<td>Retrospective, dynamic data delivery at record level</td>
</tr>
<tr>
<td>Data Warehousing &amp; Decision Support</td>
<td>What were unit sales in New England last March? Drill down to Boston?</td>
<td>Online analytic processing (OLAP), multidimensional databases, data warehouses</td>
<td>Retrospective, dynamic data delivery at multiple levels</td>
</tr>
<tr>
<td>Intelligent Data Mining</td>
<td>What’s likely to happen to Boston unit sales next month? Why?</td>
<td>Advanced algorithms, multiprocessor computers, massive databases</td>
<td>Prospective, proactive information delivery</td>
</tr>
</tbody>
</table>

SOURCE: Courtesy of Pilot Software, Inc.

**DATA COLLECTION AND ACCESS**

The techniques used for data collection and access are well known and are described in Appendix C to this text. The problem with such techniques is that they are not suitable for a large volume of data, nor can they be used effectively by end users. Even though structured query language (SQL) use is becoming more user friendly, supported by natural language processors, business questions that can be answered with it are limited.

**ONLINE ANALYTICAL PROCESSING.** Online Analytical Processing (OLAP) refers to such end-user activities as DSS modeling using spreadsheets and graphics, which are done on line. Unlike online transaction processing (OLTP) applications, OLAP involves many data items (frequently many thousands or even millions) in complex relationships. One objective of OLAP is to analyze these relationships and look for patterns, trends, and exceptions.

An OLAP database may consist of sales data that have been aggregated by region, product type, and sales channel. A typical OLAP query might access a
Access very large amounts of data—several years of sales data

- Analyze the relationships between many types of business elements, such as sales, products, regions, and channels
- Involve aggregated data, such as sales volumes, budgeted dollars, and dollars spent
- Compare aggregated data over hierarchical time periods—monthly, quarterly, yearly
- Present data in different perspectives, such as sales by region versus sales by channel or by product within each region
- Involve complex calculations between data elements, such as expected profit as calculated as a function of sales revenue for each type of sales channel in a particular region
- Are able to respond quickly to user requests so that users can pursue an analytical thought process without being stymied by the system

Here is an example of how OLAP works with a query. An end user makes a request, possibly in a natural language. Some vendors, such as SAS Institute, Inc., provide natural language interfaces (such as SAS English) for this purpose. The query is transformed into an SQL format. This query is delivered via a network to a DBMS that manages the database or the data warehouse. The DBMS finds the answer to the query and delivers it back to the end user. The user can then design a presentation or a report to fit his or her display needs.

Although OLAP is very useful in many cases, it is retrospective in nature and cannot provide the automatic and prospective knowledge discovery, which is done by advanced data mining techniques.

**DATA MINING.** Data mining derives its name from the similarities between searching for valuable business information in a large database—for example, finding linked products in gigabytes of store scanner data—and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

- **Automated prediction of trends and behaviors.** Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly and quickly from the data. A typical example of a predictive problem is targeted marketing. Data mining can use data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default and identifying segments of a population likely to respond similarly to given events.

- **Automated discovery of previously unknown patterns.** Data mining tools identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together, such as baby diapers and beer. Other pat-
tern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that may represent data entry keying errors.

When data mining tools are implemented on high-performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

Data mining can be conducted by nonprogrammers and it appears under different names, such as knowledge extraction, data dipping, data archeology, data exploration, data pattern processing, data dredging, and information harvesting. The following are the major characteristics and objectives of data mining:

1. Data are often buried deep within very large databases, such as data warehouses, that sometimes contain data stored for several years.
2. In some cases the data are consolidated in a data warehouse and data marts; in others they are kept in Internet and intranet servers.
3. The data mining environment usually has a client–server architecture.
4. Data mining tools help remove the information “ore” buried in corporate files or archived public records.
5. The “miner” is often an end user, empowered by “data drills” and other power query tools to ask ad hoc questions and get answers quickly, with little or no programming skill.
6. “Striking it rich” often involves finding unexpected, valuable results.
7. Data mining tools are easily combined with spreadsheets and other end-user software development tools; therefore, the mined data can be analyzed and processed quickly and easily.
8. Because of the large amounts of data, it is sometimes necessary to use parallel processing for data mining.
9. Data mining yields five types of information: (a) association, (b) sequences, (c) classifications, (d) clusters, and (e) forecasting.
10. Data miners can use one or several tools and techniques (see the list in Box 10.2).

A Closer Look  BOX 10.2
DATA MINING TECHNIQUES

The most commonly used techniques for data mining are:

**Case-based reasoning.** The case-based reasoning approach uses historical cases to recognize patterns. For example, customers of Cognitive Systems, Inc. utilize such an approach for help desk applications. One company has a 50,000 query case library. New cases are matched quickly against the 50,000 samples in the library, providing more than 90 percent accurate and automatic answers to queries.

**Neural computing.** Neural computing is a machine learning approach by which historical data can be examined for pattern recognition. These patterns can then be used for making predictions and for decision support (details are given in Chapter 11). Users equipped with neural computing tools can go through huge databases and, for example, identify potential customers of a new product or companies whose profiles suggest that they are heading for bankruptcy. Most current applications are in financial services (Trippi and Turban, 1996), in marketing and in manufacturing.

**Intelligent agents.** One of the most promising approaches to retrieving information from the Internet or from intranet-based databases is the use of intelligent agents. As vast amounts of information become available through the Internet, finding the right information is more difficult. This topic is discussed further in Chapters 5 and 11.

**Other tools.** Several other tools can be used. These include decision trees, genetic algorithms, nearest neighbor method, and rule reduction. For details, see Inmon (1996).
A Sampler of Data Mining Applications. Data mining can be very helpful as shown by the representative examples that follow. Note that the intent of most of these examples is to identify a business opportunity to create a sustainable competitive advantage.

1. Retailing and sales. Predicting sales; determining correct inventory levels and distribution schedules among outlets.

2. Banking. Forecasting levels of bad loans and fraudulent credit card use, credit card spending by new customers, and which kinds of customers will best respond to (and qualify for) new loan offers.

3. Manufacturing and production. Predicting machinery failures and finding key factors that control optimization of manufacturing capacity.

4. Brokerage and securities trading. Predicting when bond prices will change; forecasting the range of stock fluctuations for particular issues and the overall market; determining when to trade stocks.

5. Insurance. Forecasting claim amounts and medical coverage costs; classifying the most important elements that affect medical coverage; predicting which customers will buy new policies.

6. Computer hardware and software. Predicting disk-drive failures; forecasting how long it will take to create new chips; predicting potential security violations.

7. Policework. Tracking crime patterns, locations, and criminal behavior and attributes to assist in solving criminal cases.

8. Government and defense. Forecasting the cost of moving military equipment; testing strategies for potential military engagements; predicting resource consumption.

9. Airlines. Capturing data on where customers are flying and the ultimate destination of passengers who change carriers in midflight; thus, airlines can identify popular locations that they do not service and check the feasibility of adding routes to capture lost business.

10. Health care. Correlating demographics of patients with critical illnesses. Using data mining, doctors can develop better insights on symptoms and their causes and how to provide proper treatments.

11. Broadcasting. Predicting what is best to show during prime time and how to maximize returns by interjecting advertisements.

12. Marketing. Classifying customer demographics that can be used to predict which customers will respond to a mailing or buy a particular product, as illustrated by Marriott Club International.

---

**IT at Work:**

**DATA MINING AT MARRIOTT FOR TARGETED ADVERTISING**

Marriott Club International, the nation’s largest seller of vacation time-share condos, had a problem. The company had a database with millions of names. Marriott used to send advertisements to all of the customers in the database, at a great expense, but the response was minimal. The company decided to identify the customers in their list who were more likely to respond. By doing so, they have slashed the amount of mail considerably and improved the response rate.

Marriott uses neural computing technology in their data mining, the objective of which is to detect patterns by combing through the digitized customer lists.

Marriott started with names, mostly hotel guests. Digging into a trove of motor vehicle records, property records, warranty cards, and lists of people who had bought by mail, a computer program enriched the prospect list. It added such facts as the customers’ ages, their children’s ages, their estimated income, what cars
they drove, and whether they played golf. The Marriott system then identifies who is most likely to respond to a mailed flier.

Using these clues, Marriott has been able to cast its net a little more narrowly and catch more fish. Data mining has increased the response rate to Marriott’s direct mail time-share pitches to certain hotel guests by 33 percent, for a significant saving.

FOR FURTHER EXPLORATION. How can the detected patterns affect the response rate? Also, like many other companies, Marriott is using information on customers without their knowledge, invading their privacy. What can the company do to act more ethically? 


10.3 DATA VISUALIZATION TECHNOLOGIES

Once data have been processed they can be presented to users as text, as tables, and via several data visualization technologies.

DATA VISUALIZATION

Visual technologies make pictures worth a thousand numbers and make IT applications more attractive and understandable to users. Data visualization is achieved by technologies that support visualization of information. These include digital images, geographical information systems, graphical user interfaces, multidimensional tables and graphs, virtual reality, three-dimensional presentations, and animation. Visualization software packages offer users capabilities for self-guided exploration and visual analysis of large amounts of data. By using visual analysis technologies, people may spot problems that have existed for years undetected by standard analysis methods. Visualization technologies can also be integrated among themselves to create a variety of presentations, as Haworth Corporation has discovered.

IT at Work:
DATA VISUALIZATION HELPS HAWORTH TO COMPETE

Manufacturing office furniture is an extremely competitive business. Haworth Corporation operates in this environment and has been able to survive and even excel with the help of IT. To compete, Haworth allows its customers to customize what they buy. But it may surprise you to learn that an office chair can be assembled in 200 ways. The customization of all products resulted in 21 million potential products, confusing customers who could not know what they had bought until the item was delivered.

The solution was computer visualization software that allowed sales representatives with laptops to show customers exactly what they were ordering. Thus, the huge parts catalogs became more easily understood, and sales representatives were able to configure different options by entering the corporate database, showing what a product would look like, and computing its price.

The customers can now make changes until the furniture design meets their needs. The salesperson can do all this from the customer’s office by connecting to the corporate intranet via the Internet and using Web tools to allow customers to make the desired changes.

The program allows the company to reduce cycle time. After the last computer-assisted design (CAD) mockup of an order has been approved, the CAD software is used to create a bill of materials that goes to Haworth’s factory for manufacture. This reduces the time spent between sales reps and CAD operators, increasing the time available for sales calls and increasing customer satisfaction with quicker delivery.

By using this visualization computer program, Haworth has increased its competitive advantage in three ways:
Data visualization is easier to implement when the necessary data are in a data warehouse. Our discussion here is focused mainly on the data visualization techniques of multidimensionality and geographical information systems. Related topics, such as multimedia and hypermedia, are presented in Appendix B.

**MULTIDIMENSIONALITY**

Modern data and information may have several dimensions. For example, management may be interested in examining sales figures in a certain city by product, by time period, by salesperson, and by store. The more dimensions involved, the more difficult it is to present multidimensional information in one table or in one graph. Therefore, it is important to provide the user with a technology that allows him or her to add, replace, or change dimensions quickly and easily in a table and/or graphical presentation. Such changes are known as “slicing and dicing” of data. The technology of slicing, dicing, and similar manipulations is called multidimensionality.

Figure 10.5 shows three views of the same data, organized in different ways, using multidimensional software, usually available with spreadsheets. In part (a), travel hours of a company’s employees by means of transportation and by country are given. The “next year” gives projections automatically generated by an embedded formula. In part (b) the data are reorganized and in part (c) they are reorganized again and manipulated as well. All this is easily done by the end user with one or two clicks of the mouse.

The major advantage of multidimensionality is that data can be organized the way managers like to see them rather than the way that the system analysts do. Furthermore, different presentations of the same data can be arranged and rearranged easily and quickly. Three factors are considered in multidimensionality: dimensions, measures, and time.

- **Examples of dimensions:** Products, salespeople, market segments, business units, geographical locations, distribution channels, country, industry
- **Examples of measures:** Money, sales volume, head count, inventory profit, actual versus forecasted
- **Examples of time:** Daily, weekly, monthly, quarterly, yearly

For example, a manager may want to know the sales of product M in a certain geographical area, by a specific salesperson, during a specified month, in terms of units. Although the answer can be provided regardless of the database structure, it can be provided much faster, and by the user himself or herself, if the data are organized in *multidimensional databases* or if the query tools are designed for multidimensionality. In either case, users can navigate through the many dimensions and levels of data via tables or graphs and then conduct a quick analysis to find significant deviations or important trends.

Multidimensionality is available with different degrees of sophistication and is especially popular in executive information and support systems. There are several types of software from which multidimensional systems can be constructed, and they often work in conjunction with OLAP tools.
A geographical information system (GIS) is a computer-based system for capturing, storing, checking, integrating, manipulating, and displaying data using digitized maps. Its most distinguishing characteristic is that every record or digital object has an identified geographical location. By integrating maps with spatially oriented

Figure 10.5  Multidimensionality views.

**GEOGRAPHICAL INFORMATION SYSTEMS (GIS)**
(geographical location) databases (called “geocoding”) and other databases, users can generate information for planning, problem solving, and decision making, increasing their productivity and the quality of their decisions, as many banks have done.

**IT at Work:**

**BANKS USING GEOGRAPHICAL INFORMATION SYSTEMS (GIS) TO SUPPORT MARKETING**

Banks are using GIS for plotting the following:

- Branch and ATM locations
- Customer demographics (e.g., residence, age, income level) for each product of the bank
- Volume and traffic patterns of business activities
- Geographical area served by each branch
- Market potential for banking activities
- Strengths and weaknesses against the competition
- Branch performance

A GIS is used as a geographical spreadsheet that allows managers to model business activities and perform “what-if” analyses (What if we close a branch or merge branches? What if a competitor opens a branch?). The maps consolidate pages of analysis. Representative pioneering banks are First Florida Bank (Tampa, FL), Marion Bank (Philadelphia, PA), and NJB Financial (Princeton, NJ).

**FOR FURTHER EXPLORATION.** How can a GIS indicate a bank’s strengths and weaknesses against the competition?

---

**GIS SOFTWARE.** GIS software varies in its capabilities from simple computerized mapping systems to enterprise-wide tools for decision support data analysis (see Minicase 1). Clearly, a high-quality graphics display and high computation and search speeds are necessary, so most early GIS implementations were developed for mainframes. Since the early 1990s, however, relatively powerful PC implementations have been developed.

**GIS DATA.** GIS data are available from a wide variety of sources. Government sources (via the Internet and CD-ROM) provide some data, while vendors provide diversified commercial data as well (such as CD-ROMs from MapInfo and FirstMap from Wessex Inc.). The field of GIS can be divided into two major categories: func-
GIS AND DECISION MAKING. Initially, the high cost of GIS prevented its use outside experimental facilities and government agencies. Since the 1990s, however, the cost of GIS software and its required hardware has dropped dramatically. Now relatively inexpensive, fully functional PC-based packages are readily available. GIS provides a large amount of extremely useful information that can be analyzed and utilized in decision making. Its graphical format makes it easy for managers to visualize the data. For example, explains Janet M. Hamilton, market research administrator for Dow Elanco, a $2 billion maker of agricultural chemicals based in Indianapolis,

I can put 80-page spreadsheets with thousands of rows into a single map. It would take a couple of weeks to comprehend all of the information from the spreadsheet, but in a map, the story can be told in seconds. (Hamilton, 1996, p. 21)

There are countless applications of GIS to better decision making in the public or private sector. They include the dispatch of emergency vehicles, transit management (see Minicase 1), facility site selection, and wildlife management.

For many companies, the intelligent organization of data within a GIS can provide a framework to support the process of decision making and of designing alternative strategies. Some examples of successful GIS applications are provided by Bidgoli (1995) and Hamilton (1996). Some examples of successful GIS application are summarized in A Closer Look 10.3.

GIS AND THE INTERNET OR INTRANET. Most major GIS software vendors are providing Web access, such as embedded browsers, or a Web/Internet/intranet server that hooks directly into their software. Thus users can access dynamic maps and data via the Internet or a corporate intranet (see Jacobs, 1996). Big Horn Computer Services, in Buffalo, WY, is using a Web-adapted GIS to develop a custom application for a national television network that wants its affiliate stations to be able to access an intranet containing demographic information about their viewers. Using a Web browser, employees at each station can view thematically shaded maps analyzing their market (see Swenson, 1996).
A number of firms are deploying GIS on the Internet for internal use or for use by their customers. For example, Visa Plus, which operates a network of automated teller machines, has developed a GIS application that lets Internet users call up a locator map for any of the company’s 257,000 ATM machines worldwide. As GIS Web server software is deployed by vendors, more applications will be developed. Maps, GIS data, and information about GIS are available over the Web through a number of vendors and public agencies.

EMERGING GIS APPLICATIONS. The integration of GIS and global positioning systems (GPS) has the potential to help reengineer the aviation and shipping industries. It enables vehicles or aircraft equipped with a GPS receiver to pinpoint their location as they move (Tetzeli, 1993). Emerging applications of GPS include personal automobile mapping systems, railroad car tracking, and earth-moving equipment tracking. Now the price is rather high, but it is dropping with improvements in hardware, increased demand, and the availability of more competing vendors. GPS has also become a major source of new GIS data (see Group Assignment 1).

Some researchers have developed intelligent GISs that link a GIS to an expert system (such as applications in irrigation scheduling and forest pest management; see Plant, 1993).

Improvements in the GIS user interface have substantially altered the GIS “look and feel.” Advanced visualization (three-dimensional graphics) is increasingly integrated with GIS capabilities, especially in animated and interactive maps. GIS can provide information for virtual reality engines and it can display complex information to decision makers. Object-oriented programming and databases are also likely to improve GIS. Multimedia and hypermedia will play a growing role in GIS, especially in help and training systems. Object linking and embedding will allow users to import maps into any document. More GIS will be deployed to provide data and access data over the Web and organizational
intranets as “Web-ready” GIS software becomes available. See Clarke (1997) for an overview of GIS, many details, recent developments, and potential advances.

### 10.4 MARKETING DATABASES IN ACTION

Data warehouses and data marts serve end users in all functional areas. However, the most dramatic applications of data warehousing and mining are in marketing.

In this section we examine how data warehouses, their extensions, and data mining are used and what role they play in new marketing strategies, such as the use of marketing transaction databases in interactive marketing.

#### THE MARKETING TRANSACTION DATABASE*

Success in interactive marketing requires a new kind of database, oriented toward targeting and personalizing marketing messages in real time. Such a database provides the most effective means of capturing information on customer preferences and needs and, in turn, enterprises can use this knowledge to create new products and services. Such a database is called a **marketing transaction database (MTD)**. Most current databases are static: they simply gather and store information about customers. They fall under the following categories: operations databases, data warehouses, and marketing databases. The MTD combines many of the characteristics of these databases and marketing data sources into a new database that allows marketers to engage in real-time personalization and target every interaction with customers.

**MTD’S CAPABILITIES.** The MTD provides dynamic, or interactive, functions not available with traditional types of marketing databases. In marketing terms, a transaction occurs with the exchange of information. With interactive media, each exposure to the customer becomes an opportunity to conduct a marketing “transaction.” Exchanging information (whether gathered actively, through registration or use requests, or passively, by monitoring customer behavior) allows marketers to refine their understanding of each customer continuously and to use that information to target him or her specifically with personalized marketing messages.

A comparison of MTDs and other marketing-related databases is given in Table 10.3. Note that MTDs do not eliminate the traditional databases. They complement them by providing additional capabilities.

**THE ROLE OF THE INTERNET.** Data mining, data warehousing, and MTDs are delivered on the Internet and intranets. The Internet does not simply represent another advertising venue or a different medium for catalog sales, it contains new attributes that smart marketers can exploit to their fullest degree. Indeed, the Internet promises to revolutionize sales and marketing. Dell Computers (see Minicase 2) offers a good example of how marketing professionals can use the Internet’s electronic sales and marketing channels for market research, advertising, information dissemination, product management, and product delivery.

#### DETAILED IMPLEMENTATION EXAMPLES

Fewer and fewer companies can afford traditional marketing approaches, which include big-picture strategies and expensive marketing campaigns. Marketing departments are being scaled down and new approaches such as one-to-one mar-

marketing, speed marketing, interactive marketing, and relationship marketing are being employed. The following examples illustrate how companies use data mining and warehousing to support the new marketing approaches.

- Alamo Rent-a-Car discovered that German tourists liked bigger cars. So now, when Alamo advertises its rental business in Germany, the ads include information about its larger models.

- Burlington Coat Factory tracks families as they grow through the online registry for expectant parents. The company then matches direct-mail material to the different stages of a family’s development over time. Burlington also identifies, on a daily basis, top-selling styles and brands. By digging into reams of demographic data, historical buying patterns, and sales trends in existing stores, Burlington determines where to open its next store and what to stock in each store. The company’s data warehouse contains 1.5 terabytes of information.

- Au Bon Pain Company Inc., a Boston-based chain of cafes, discovered that the company was not selling as much cream cheese as planned. When it analyzed point-of-sale data, the firm found that customers preferred small, one-serving packaging (like butter). As soon as the package size was changed, sales shot up.

- Using U.S. census data along with its own internal data, Spalding Sports profiled thousands of golf courses and pro shops throughout the country. Promotional materials for each golf course now match the customers’ profile (such as upscale golfers versus working class tourists). They also found that buyers at pro shops were more interested in technical aspects than buyers at retail stores.

- Bank of America gets more than 100,000 telephone calls from customers every

---

**Table 10.3 Marketing Databases**

<table>
<thead>
<tr>
<th></th>
<th><strong>Operational System</strong></th>
<th><strong>Data Warehouse</strong></th>
<th><strong>Data Marts Marketing Analysis Database</strong></th>
<th><strong>Marketing Transaction Database</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Record transactions</td>
<td>Decision making</td>
<td>Understanding behavior</td>
<td>Targeting and personalization</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>Rigid—tied to applications and functional rules</td>
<td>Loose—iterative, modifiable</td>
<td>Loose, but marketing specific</td>
<td>Liquid—driven by real-time marketing</td>
</tr>
<tr>
<td><strong>Update frequency</strong></td>
<td>Real time</td>
<td>Weekly/monthly</td>
<td>Monthly/quarterly</td>
<td>Real time</td>
</tr>
<tr>
<td><strong>Data level</strong></td>
<td>Transaction detail</td>
<td>Most summarized, some detail</td>
<td>Summarized</td>
<td>Individual detail</td>
</tr>
<tr>
<td><strong>Data quality</strong></td>
<td>Dubious—most legacy systems have poor QA</td>
<td>Rationalized (cleansed, verified)</td>
<td>Rationalized/householdered</td>
<td>Focused customer data—verified by individual</td>
</tr>
<tr>
<td><strong>Data type</strong></td>
<td>Behavioral—based on transactions</td>
<td>Descriptive, behavioral</td>
<td>Descriptive, behavioral, derivative</td>
<td>Combines derivative, behavioral, and descriptive</td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td>Very fast—high transaction volume</td>
<td>Centralized, easy to access repository</td>
<td>Organized by marketers for marketing purposes</td>
<td>Allows real-time analysis and decision making</td>
</tr>
<tr>
<td><strong>Issues</strong></td>
<td>Not oriented or organized the way marketers think</td>
<td>Difficult to implement correctly—typically needs more information for marketing</td>
<td>Generally oriented toward analysis only—little capability for real-time decision making</td>
<td>Emerging—not standards or best practices, not integrated with other systems</td>
</tr>
<tr>
<td><strong>Event trigger</strong></td>
<td>External</td>
<td>Internal</td>
<td>Internal</td>
<td>External</td>
</tr>
</tbody>
</table>

day. Analyzing their banking activities, the bank determines what may be of interest to the customers. So when a customer calls to check on a balance the bank tries to sell the customer something in which he or she might be interested. The customer database used for this contains about 1 terabyte of data, providing insight into customers’ behaviors.

✦ Supermarket chains regularly analyze reams of cash register data to discover what items customers are typically buying at the same time. These shopping patterns are used for issuing coupons, designing floor layouts (diapers are placed near to beer; can you guess why?), and creating shelf displays.

✦ AT&T and MCI sift through terabytes of customer phone data to fine tune marketing campaigns and determine new discount calling plans.

✦ A pharmaceutical company analyzes the results of its recent sales force activity to improve targeting of physicians who should be first contacts and determine which marketing activities will have the greatest impact in the next few months. The data include competitor market activity as well as information about the local health care systems. The results are distributed to the sales force via the Internet, intranets, or a private wide-area network that enables the representatives to review the decision-making recommendations.

✦ A credit card company leverages its vast warehouse of customer transaction

![Netscape: The Chicago Tribune](image)

**FIGURE 10.7** The Chicago Tribune’s web page.
data to identify customers most likely to be interested in a new credit product. Using a small test mailing, it identifies the attributes of customers with an affinity for the product. Recent projects have indicated more than a 20-fold decrease in costs for targeted mailing campaigns over conventional approaches.

- A diversified transportation company with a large direct sales force applies data mining to identifying the best prospects for its services. Using data mining to analyze its own customer experience, this company can build a unique segmentation identifying the attributes of high-value prospects. Applying this segmentation to a general business database, such as those provided by Dun & Bradstreet, can yield a prioritized list of prospects by region.

- A large consumer package goods company applies data mining to improve its sales process to retailers. Data from consumer panels, shipments, and competitor’s activity are examined to understand the reasons for brand and store switching. Through this analysis, the manufacturer can select promotional strategies that best reach its target customer segments.

- In its data warehouse, the Chicago Tribune stores information about customer behavior as customers move through the various newspaper Web sites. Data mining helps to analyze volumes of data ranging from what browsers are used to what hyperlinks are clicked on most frequently (see Figure 10.7).

The data warehouses in some companies include 1 terabyte or more of data. They need to use supercomputing to sift quickly through the data. Wal-Mart, the world’s largest discount retailer, has a gigantic database.

---

**IT at Work:**

DATA MINING POWERS WAL-MART

With more than 7 terabytes of data on two NCR (National Cash Register) systems, Wal-Mart manages one of the world’s largest data warehouses. Besides the two NCR Teradata databases, which handle most decision-support applications, Wal-Mart has another 6 terabytes of transaction processing data on IBM and Hitachi mainframes.

Wal-Mart’s formula for success—getting the right product on the appropriate shelf at the lowest price—owes much to the company’s multimillion dollar investment in data warehousing. “Wal-Mart can be more detailed than most of its competitors on what’s going on by product, by store, by day—and act on it,” says Richard Winter, a database consultant in Boston. “That’s a tremendously powerful thing.”

The systems house data on point of sale, inventory, products in transit, market statistics, customer demographics, finance, product returns, and supplier performance. The data are used for three broad areas of decision support: analyzing trends, managing inventory, and understanding customers. What emerges are “personality traits” for each of Wal-Mart’s 3000 or so outlets, which Wal-Mart managers can use to determine product mix and inventory levels for each store.

Wal-Mart is rolling out a data mining demand-forecasting application based on neural networking software and a 4000-processor parallel computer. The application “looks at individual items for individual stores to decide the seasonal sales profile of each item,” says Rick Dalzell, Wal-Mart’s Vice President of Applications Development. The computer keeps a year’s worth of data on the sales of 100,000 products and predicts which items will be needed in each store and when.

Wal-Mart is expanding its use of market basket analysis. Data are collected on items that comprise a shopper’s total purchase so that the company can analyze relationships and patterns in customer purchases. The data warehouse is available over the Web to store managers and suppliers. In 1997, 3500 users made 20,000 database queries each day.

“What Wal-Mart is doing is letting an army of people use the database to make tactical decisions,” says consultant Winter. “The cumulative impact is immense.”

**FOR FURTHER EXPLORATION.** Why is data dipping necessary? What kind of competitive advantage can Wal-Mart gain with the system? ▲

10.5 KNOWLEDGE MANAGEMENT

Knowledge Bases and Organizational Learning

Throughout this chapter we have shown how raw data are transformed to information and knowledge. We have demonstrated how knowledge is derived for problem solving and how it is used for forecasting or predicting customer behavior. It only makes sense that the knowledge created for solving problems will be reused whenever the organization faces the same or similar problems (why reinvent the wheel?). This simple idea is the basis for a challenging concept called knowledge management or managing knowledge bases. A knowledge base is a database that contains knowledge, or the organization’s knowhow. The major purpose of an organizational knowledge base is to allow for knowledge sharing (see Manville and Foote, 1996). Knowledge sharing among employees, with customers, and with business partners has a huge potential payoff in improved customer service, shorter delivery cycle times, and increased collaboration within the company and with business partners as well. Furthermore, some knowledge can be sold to others or traded for other knowledge.

Sharing knowledge also means facilitating organizational learning. People can learn from the experience of others when this experience is documented in the knowledge base.

Implementing Knowledge Management Systems

Building and maintaining organizational knowledge bases is not a simple task and it involves many activities. For example, Arthur Andersen, one of the largest CPA companies, is pioneering an organizational knowledge base.

IT at Work: Arthur Andersen’s Knowledge Base—Leapfrogging the Competitors

For a large consulting company, whose very product is knowledge, there is considerable motivation to create a knowledge base to share accumulated knowhow. For this reason, Arthur Andersen and its sister company, Andersen Consulting, began in the early 1990s to create their Global Best Practices (GBP) knowledge base, a central repository of knowledge about world-class business practices.

The GBP contains quantitative and qualitative information about how companies achieve best-in-the-world standards of performance in activities that are common to most companies. Andersen’s consultants use this information to provide clients with an ever-growing body of knowledge that can be used for performance improvement. Providing such knowledge to its customers gives a competitive edge to the company. Four years after the GBP was created, it was so successful that it had fundamentally shifted the company’s culture and the way it did business. As a matter of fact, the company developed several other knowledge bases, each dedicated to an important topic. Here is how Arthur Andersen gathers knowledge.

- The Global Best Practices hotline receives more than 250 calls each month, which are sorted into business practice categories.
- These data, combined with ongoing research into emerging areas, are used to determine which process areas are to be developed or enhanced for an upcoming release.
- Research analysts team with content experts, who serve clients, to develop best practices context and diagnostic tools.
- Qualitative and quantitative information and tools are released on a CD-ROM that can be accessed by all Arthur Andersen professionals worldwide.
- Use in the field with clients generates suggestions that are received by the hotline or captured through formal or informal surveys.
From the Andersen example, one can see that implementing a knowledge base is not a simple project. The cost of building and maintaining it can be very high and the benefits are difficult to justify. Some implementation issues are: Who will decide on what to include in the knowledge base and how? Who will extract the knowledge from its sources? How will managers and professionals be trained to make effective use of the knowledge? What portion of the knowledge base should be open to outsiders, and how can proprietary and sensitive information be secured? How will the firm integrate the knowledge base with existing databases? And finally, how will managers validate the quality of the knowledge? Despite these issues, companies are considering not only creating knowledge bases, but also completely reorganizing themselves as knowledge-based organizations (see Chapter 19 in Holsapple and Whinston, 1996). Even a new position, titled chief knowledge officer, or CKO, has been created (Davenport, 1996), whose role is to capture and leverage structured knowledge with IT. CKOs have three responsibilities: creating knowledge management infrastructure, building a knowledge culture, and making it all pay off. A related position is the chief learning officer (established in the late 1990s at Coca-Cola and General Electric).

Appointing a CKO may be the first step toward fulfilling Drucker’s (1988) prediction that “the typical business will be knowledge based, an organization composed largely of specialists who direct and discipline their own performance through organized feedback from colleagues, customers, and headquarters”. Drucker claims that to remain competitive, or maybe even to survive, businesses will have to convert themselves into organizations of knowledge specialists. Based on Drucker’s vision, Leonard-Barton (1995) and Holsapple and Whinston (1996) have developed frameworks for the knowledge organizations of tomorrow.

**FOR FURTHER EXPLORATION.** Can you think of other topics that might be the subject of a knowledge base in a consulting company? In a university?

The knowledge base contains the following information:
- Best company profiles
- Relevant Arthur Andersen engagement experience
- Top 10 case studies and articles
- World-class performance measures
- Diagnostic tools
- Customizable presentations
- Process definitions and directory of internal experts
- Best control practice
- Tax implications

**MANAGERIAL ISSUES**

A number of challenges face companies that are undertaking a data and knowledge management approach.

**Cost–benefit issues and justification.** Some of the solutions discussed here are very expensive and are justifiable only in large corporations. Smaller organizations can make the solutions cost effective if they leverage existing databases rather than create new ones. A careful cost–benefit analysis must be undertaken before any commitment to the new technologies is made.

**Where to store data physically.** Should data be distributed close to their sources, thereby potentially speeding up data entry and updating but increasing problems of data security and backup, or should data be centralized for easier control, security, backup, and disaster recovery, although they will be more distant from users and create a potential single point-of-failure location?
Disaster recovery. Can an organization’s business processes that have become dependent on databases recover and sustain operations after a natural or other type of information system disaster?

Internal or external? Should a firm invest in internally collecting, storing, maintaining, and purging its own databases of information, or should it pay to subscribe to external databases, where providers are responsible for all data management and data access?

Data security and ethics. Are the company’s competitive data safe from external snooping or sabotage? Are confidential data, such as personnel details, safe from improper or illegal access and alteration?

Data purging. When is it beneficial to “clean house” and purge information systems of obsolete or non–cost effective data?

The legacy data problem. One very real problem is what to do with the mass of information already stored in a variety of formats, often known as the legacy data acquisition problem. Data in older, perhaps obsolete, databases still need to be available to newer database management systems. Many of the legacy application programs used to access the older data simply cannot be converted into new computing environments without both transparent and procedural access to critical data remaining in the legacy environment. Basically, there are three approaches to solving this problem. One is to create a database front end that can act as a translator from the old system to the new. The second is to cause applications to migrate into the new system, so that data can be seamlessly accessed in the original format. The third is to cause the data to migrate into the new system by reformattting it.

Data delivery. A problem regarding how to move data efficiently around an enterprise also exists. The inability to communicate among different groups in different geographical locations is a serious roadblock to implementing distributed applications properly, especially given the many remote sites and mobility of today’s workers.

Privacy. Collecting data in a warehouse and conducting data mining may result in the invasion of individual privacy. What will companies do to protect the individuals? What can the individuals do to protect their privacy?
The Internet is a major source of data and knowledge. (L-2)

Many factors that impact the quality of data must be recognized and controlled. (L-2)

The newest types of data organization are object oriented and multimedia. (L-3)

Electronic document management, the automated control of documents, is a key to greater efficiency in handling documents in order to gain an edge on the competition. (L-4)

Warehouses and data marts are necessary to support effective decision making. Relevant data are indexed and organized for easy access by end users. (L-4)

Data mining for knowledge discovery is an attempt to use intelligent systems to scan volumes of data to locate necessary information and knowledge. (L-5)

Online analytical processing is a data discovery method that uses analytical approaches to knowledge discovery. (L-5)

Visualization is important for better understanding of data relationships and compression of information. (L-6)

Multidimensional presentation enables quick and easy multiple viewing of information in accordance with people’s needs. (L-6)

A geographical information system captures, stores, manipulates, and displays data using digitized maps. (L-6)

Database marketing provides the technological support for new marketing approaches such as interactive marketing. (L-7)

Marketing transaction databases provide dynamic interactive functions that facilitate customized advertisement and services to customers. (L-7)

Knowledge bases contain shareable knowledge accumulated in organizations. Their management is difficult, but their use can be extremely beneficial. (L-8)

QUESTIONS FOR REVIEW

1. Define knowledge management.
2. Define marketing transaction database.
3. Define knowledge-based organization.
4. What is a terabyte? (Write the number.)
5. Review the steps of the data life cycle and explain them.
6. List some of the categories of data available on the Internet.
7. Define data quality.
8. Define document management.
10. Define a data mart.
11. Define online analytical processing (OLAP).
12. Define data mining.
13. Describe the properties of multifunctionality.
15. Define marketing transaction database.
16. Define knowledge base and describe its major benefits.

QUESTIONS FOR DISCUSSION

1. Relate knowledge management to organizational learning.
2. Discuss the opportunities that a knowledge base can provide to a pharmaceutical company such as Merck or Johnson & Johnson.
3. Compare data quality to data integrity. How are they related?
4. Discuss some of the advantages of object-oriented databases.
5. Discuss the factors that make document management so valuable. What capabilities are particularly valuable?
6. Relate document management to imaging systems.
7. Describe the process of knowledge discovery and discuss the roles of the data warehouse, data mining, and OLAP in this process.
8. Discuss the major benefits of data warehousing to end users.
9. Discuss how a data warehouse can lessen the stovepipe problem.
10. A data mart can substitute for a data warehouse or supplement it. Compare and discuss.
11. Why is the combination of GIS and GPS becoming so popular? Examine some applications.
12. Discuss the advantages of terabyte marketing databases to a large corporation. Does a small company...
need a marketing database? Under what circumstances will it make sense to have one?

GROUP ASSIGNMENTS

1. Several applications now combine GIS and GPS:
   a. Survey such applications by conducting literature search and query GIS vendors.
   b. Prepare a list of five applications.
   c. Describe the benefit of such integration.

2. Prepare a report on the topic of “data management and the intranet.” Specifically pay attention to delivery of the data warehouse, use of browsers for query, and data mining. Also explore the issue of GIS and the Internet. Finally, describe the role of extranets in support of business partner collaboration. Each student will visit one or two vendors’ sites, read the white papers, and examine products (Oracle, Red Bricks, Brio, Siemens Mixdorf IS, Comshare, NCR, SAS, and Information Advantage). Also, visit the Web site of the Data Warehouse Institute.

EXERCISES

1. Review the list of data management difficulties in Section 10.1. Explain how a combination of a data warehousing and data mining can solve or reduce these problems. Be specific.

2. Interview a knowledge worker in a company you work for or to which you have access. Find the data problems they have encountered and the measures they have taken to solve them. Relate the problems to Strong, et al.’s four categories.

3. Prepare a report about knowledge management. Use terms such as “learning organizations,” “corporate memory,” and “intellectual capital” in your reports. Visit http://world.stl.com/Lo to discern the ideas driving learning organizations.

4. Ocean Spray Cranberries is a large cooperative of fruit growers and processors. Ocean Spray needed data to determine the effectiveness of its promotions and its advertisements and to make itself able to respond strategically to its competitors’ promotions. The company also wanted to identify trends in consumer preferences for new products and to pinpoint marketing factors that might be causing changes in the selling levels of certain brands and markets. Ocean Spray buys marketing data from InfoScan, a company that collects data using bar code scanners in a sample of 2500 stores nationwide. The data for each product include sales volume, market share, distribution, price information, and information about promotions (sales, advertisements). The amount of data provided to Ocean Spray on a daily basis is overwhelming (about 100 to 1000 times more data items than Ocean Spray used to collect on its own).

   a. Find information about InfoScan and Ocean Spray by entering Ocean Spray’s Web site.
   b. Ocean Spray has said that it cannot run the business without the system. Why?
   c. What data from the data mart are used by the DSS?

INTERNET EXERCISES


2. Access the Web sites of one or two of the major data management vendors, such as Oracle, Informix, and Sybase, and trace the capabilities of their latest products, including Web connections.

13. Discuss the difficulties associated with creating and maintaining knowledge bases.

3. Companies invest billions of dollars to support database marketing. The information systems departments’ (ISD) activities that have supported accounting and finance in the past are shifting to marketing. According to Tucker (1997), some people think that the ISD should report to marketing. Do you agree or disagree? Debate this issue.

4. In 1996 Lexis-Nexis, the online information service, was accused of permitting access to sensitive information on individuals. Using data mining, it is possible not only to capture information that has been buried in distant courthouses, but also to manipulate and crossindex it. This can benefit law enforcement but invade privacy. The company argued that the firm was targeted unfairly, since it only provided basic residential data for lawyers and law enforcement personnel. Should Lexis-Nexis be prohibited from allowing access to such information or not? Debate the issue.
3. Access the Web sites of one or two of the major data warehouse vendors, such as NCR, SAS, or Comshare; find how their products are related to the Web.

4. Access the Web site of the GartnerGroup (www.gartnergroup.com). Examine some of their research notes pertaining to marketing databases, data warehousing, and data management. Prepare a report regarding the state of the art.


6. Survey some GIS resources such as www.geo.ed.ac.uk/home/hiswww.html and www.prenhall.com/stratgis/sites.html. Identify GIS resources related to your industry and prepare a report on some recent developments or applications. See http://nsdi.usgs.gov/nsdi/pages/what_is_gis.html.

7. Visit the sites of some GIS vendors (such as MAP Info Systems). Join a newsgroup and discuss new applications in marketing, banking, and transportation. Download a demo. What are some of the most important capabilities and applications?

### Minicase 1

Public transportation in Dallas and its neighboring communities is provided by Dallas Area Rapid Transit (DART), which operates buses, vans, and a light-rail system. The service area has grown very fast. By the mid-1980s, the agency was no longer able to respond properly to customer requests, make rapid changes in scheduling, plan properly, or manage security. The solution to these problems was discovered in GIS. A GIS digitizes maps and maplike information, integrates it with other database information, and uses the combined information for planning, problem solving, and decision making. DART maintains a centralized graphical database of every object for which it is responsible.

The GIS presentation makes it possible for DART's managers, consultants, and customers to view and analyze data on digitized maps. Previously, DART manually created service maps showing bus routes and schedules. The maps were updated and redistributed several times a year, at a high cost. The manual method made it difficult to respond quickly and accurately to the nearly 6000 customer inquiries each day. For example, to answer a question concerning one of the more than 200 bus routes or a specific schedule, it was often necessary to look at several maps and routes. Planning a change was also a time-consuming task. Analysis of the viability of bus route alternatives made it necessary to photocopy maps from map books, overlay tape to show proposed routes, and spend considerable time gathering information on the demographics of the corridors surrounding the proposed routes. The GIS includes attractive and accurate maps that interface with a database containing information about bus schedules, routes, bus stops (in excess of 15,000), traffic surveys, demographics, and addresses on each street in the database. The system allows DART employees to:

- Respond rapidly to customer inquiries (reducing response time by at least 33 percent)
- Perform the environmental impact studies required by the city
- Track where the buses are at any time using a global positioning system
- Improve security on buses
- Monitor subcontractors quickly and accurately
- Analyze the productivity and use of existing routes

For instance, a customer wants to know the closest bus stop and the schedule of a certain bus to take her to a certain destination. The GIS automatically generates the answer when the caller says where he or she is by giving an address, a name of an intersection, or a landmark. The computer can calculate the travel time to the desired destination as well.

Analyses that previously took days to complete are now executed in less than an hour. Special maps, which previously took up to a week to produce at a cost of $13,000 to $15,000 each, are produced in 5 minutes at the cost of 3 feet of plotter paper.

In the late 1990s, the GIS was combined with a GPS. The GPS tracks the location of the buses and computes the expected arrival time at each bus stop.

### Questions for Minicase 1

1. Describe the role of data in the DART system.
2. What are the advantages of computerized maps?
3. Comment on the following statement: "Using GIS, users can improve not only the inputting of data but also its use."
4. What information is provided by the GPS (give an example).

Dell Computer Corporation has been the world’s largest direct-sale vendor of personal computers. One way the company distinguishes itself from other suppliers of PCs is by acting quickly on the masses of data it gathers from customers (the company receives over 50,000 telephone calls or electronic mail messages daily). “Information is a valuable competitive weapon,” says Tom Thomas, the chief information officer. “Our whole business system is geared to collect it.”

Many of the 50,000 daily messages received by Dell are from potential customers who dial 800 numbers or send e-mail to reach the company’s sales representatives; the rest are from current users of Dell machines asking the technical support staff for help. The employees who take these calls work on PCs linked to a computer that contains the company’s customer database, which has well over 1 million entries. The telephone representatives enter information about each call as it occurs, recording names and addresses along with product preferences and/or technical problems. The company stores all this information and much more in a single database shared by employees in departments from marketing to product development to customer service.

The data yield significant marketing and sales guidelines. Says Tom Martin, Dell’s chief marketer, “We know that if we use a yellow background on a catalog cover, we’ll get a 30 percent lower response rate than with gray.” The company tailors its mailings even more precisely to each recipient. The rate of response to its mailings to small businesses rose 250 percent once Dell used customer feedback to refine its pitch. At a mailing price of 50¢ to $3 per piece, the benefits of accurate targeting add up quickly.

Experience from the database also guides the sales representatives who receive calls. As they enter information about each caller, sales suggestions automatically pop up on their computer screens. Dell had a 10-fold increase in sales of 3-year warranties after prompting representatives to pitch them to all callers buying systems costing more than a certain dollar amount.

Routine analysis of customer and sales data allows Dell to spot consumer trends such as a shift to larger hard disk drives. At one time, when Dell was shipping most of its systems with drives capable of storing over 120 million characters (120 MB), the customer database alerted management to the fact that new orders for drives with nearly twice the storage capability were rapidly climbing. Dell buyers rushed out, negotiated volume discounts from large disk drive manufacturers, and locked in deliveries before their competitors.

“Know your customer” is a tried and true business rule, and Dell gets the most it can out of it through customer databases.

Questions for Minicase 2

1. What role do databases play in Dell’s marketing strategies?
2. Can you identify any data mining necessary for the information described in the case?
3. Is there any possibility of invasion of privacy of Dell’s customers? If so, how can this privacy be protected?
4. The catalog business is supplemented by electronic catalogs; therefore, some suggest that the benefit of “accurate targeting” in electronic catalogs could be lessened or even disappear. Do you agree or not? Discuss.

Minicase 2

Database Marketing Increases Dell’s Sales


REFERENCES AND BIBLIOGRAPHY


PHOTO CREDITS

Chapter 6

Chapter 10
Page 96: © Chicago Tribune Company. All rights reserved. Used with permission.