CHAPTER • 5 •

Database Systems and Business Intelligence

PRINCIPLES	LEARNING OBJECTIVES
▪ Data management and modeling are key aspects of organizing data and information.	▪ Define general data management concepts and terms, highlighting the advantages of the database approach to data management. ▪ Describe the relational database model and outline its basic features.
▪ A well-designed and well-managed database is an extremely valuable tool in supporting decision making.	▪ Identify the common functions performed by all database management systems, and identify popular database management systems.
▪ The number and types of database applications will continue to evolve and yield real business benefits.	▪ Identify and briefly discuss current database applications.

Information Systems in the Global Economy Wal-Mart, United States

Warehousing and Mining Data on a Grand Scale

One company that really needs to know how to manage data is Wal-Mart. With a total of over 800 million transactions per day in over 7,000 stores around the world, Wal-Mart produces more data in a day than many businesses produce in a lifetime. No matter what size the business, databases and the systems that manage them provide the foundation on which business decisions are made.

Wal-Mart is successful due to its ability to learn from the data it collects. In a nutshell, Wal-Mart owes its success to its databases and business intelligence tools—software tools that manipulate data to provide useful information to Wal-Mart decision makers.

At Wal-Mart headquarters in Arkansas, massive amounts of data are collected every day from its stores around the world, and stored in a data warehouse over a petabyte in size—which is a quadrillion bytes or a million gigabytes. A data warehouse is a large database that collects data from many sources, which can then be analyzed to guide business decisions. Wal-Mart uses HP's Neoview technology for its data warehouse. The system integrates data warehousing hardware, software, and services to manage large amounts of data. It is ideal for a company looking for a powerful database tool that is easy to administer.

Neoview offers "next generation business intelligence" features that embed useful information mined from the data directly into the systems that executives, managers, and employees use every day. "A lot of people in the company are asking for quicker and easier access to data. We want to make sure it's readable and usable by internal customers," says Jim Scantlin, the director of enterprise information management at Wal-Mart.

Specifics about how Wal-Mart uses business intelligence are corporate secrets that the company works hard to keep from its competitors. Clearly one of the top goals of the system is to determine which products are selling well at various locations so that Wal-Mart can manage inventory and promotions. When asked about the role of business intelligence in Wal-Mart's business strategies, Wal-Mart CTO Nancy Stewart says, "Business intelligence is huge. It is huge." Without sophisticated data analyses, making decisions regarding business strategy would be like running through the woods wearing a blindfold. A data warehouse not only allows a company to navigate through current market conditions, but in many cases provides information that allows the business to predict and plan for the future.

Wal-Mart uses its databases, data warehouse, and business intelligence tools to collect, analyze, and disseminate massive amounts of data across its networks every day. Top-level executives, regional managers, store managers, and associates are provided with custom-designed reports, charts, and graphs presented in easy-to-read dashboard software that lets users understand the state of the business at any time so they can do their jobs more effectively. As a pilot watches and analyzes the gauges and meters on the control panel of a jumbo jet to provide a smooth flight, Wal-Mart executives and managers watch and analyze the dashboard of Wal-Mart's data warehouse to keep the business running smoothly.

As you read this chapter, consider the following:

- What role do databases play in the overall effectiveness of information systems?
- What techniques do businesses use to maximize the value of the information provided from databases?

Why Learn About Database Systems and Business Intelligence?

A huge amount of data is entered into computer systems every day. Where does all this data go and how is it used? How can it help you on the job? In this chapter, you will learn about database systems and business intelligence tools that can help you make the most effective use of information. If you become a marketing manager, you can access a vast store of data on existing and potential customers from surveys, their Web habits, and their past purchases. This information can help you sell products and services. If you become a corporate lawyer, you will have access to past cases and legal opinions from sophisticated legal databases. This information can help you win cases and protect your organization legally. If you become a human resource (HR) manager, you will be able to use databases and business intelligence tools to analyze the impact of raises, employee insurance benefits, and retirement contributions on long-term costs to your company. Regardless of your field of study in school, using database systems and business intelligence tools will likely be a critical part of your job. In this chapter, you will see how you can use data mining to extract valuable information to help you succeed. This chapter starts by introducing basic concepts of database management systems.

A database is an organized collection of data. Like other components of an information system, a database should help an organization achieve its goals. A database can contribute to organizational success by providing managers and decision makers with timely, accurate, and relevant information based on data. For example, at Creative Artists Agency (CAA), a successful Hollywood talent agency, a database helps agents organize information about clients.[1] With clients such as Tom Cruise, Julia Roberts, and Brad Pitt, a talent agency must prevent mistakes and misunderstandings. CAA's database can store various types of information about each client. For example, the database informs agents about movies in which Tom Cruise is acting, movies he is producing, products he is endorsing, and any other pertinent information about the actor's career. Using the database, an agent could find all clients that are associated with a particular product or film, or all the products and films associated with one client. Databases also help companies generate information to reduce costs, increase profits, track past business activities, and open new market opportunities. In some cases, organizations collaborate in creating and using international databases. Six organizations, including the Organization of Petroleum Exporting Countries (OPEC), International Energy Agency (IEA), and the United Nations, use a database to monitor the global oil supply.

A database provides an essential foundation for an organization's information and decision support system. Without a well-designed, accurate database, executives, managers, and others do not have access to the information they need to make good decisions. For example, the city of Albuquerque, New Mexico, provides its citizens with access to a database that provides information on "water bills and usage, crime statistics in specific neighborhoods, and election campaign contributions."[2] The database provides citizens with direct access to valuable information and frees city workers from having to supply the information.

A database is also the foundation of most systems development projects. If the database is not designed properly, the systems development effort can be like a house of cards, collapsing under the weight of inaccurate and inadequate data. Because data is so critical to an organization's success, many firms develop databases to help them access data more efficiently and use it more effectively. This typically requires a well-designed database management system and a knowledgeable database administrator.

database management system (DBMS)
A group of programs that manipulate the database and provide an interface between the database and the user of the database and other application programs.

database administrator (DBA)
A skilled IS professional who directs all activities related to an organization's database.

A **database management system (DBMS)** consists of a group of programs that manipulate the database and provide an interface between the database and its users and other application programs. Usually purchased from a database company, a DBMS provides a single point of management and control over data resources, which can be critical to maintaining the integrity and security of the data. A database, a DBMS, and the application programs that use the data make up a database environment. A **database administrator (DBA)** is a skilled and trained IS professional who directs all activities related to an organization's database, including providing security from intruders. A security breach at an Ivy

League college provided an intruder with access to a database that stored students' private information.[3] Such data breaches have become commonplace for businesses and organizations because many databases are now accessible from the Internet. Data quality and accuracy also continue to be important issues for DBAs. A database error in the United Kingdom left 400,000 people without paychecks in March, 2007.[4]

Databases and database management systems are becoming even more important to businesses as they deal with increasing amounts of digital information. A report from IDC, called "The Diverse and Exploding Digital Universe," estimates the size of the digital universe to be 281 exabytes, or 281 billion gigabytes. By 2011, there will be 1,800 exabytes of electronic data in existence, or 1.8 zettabytes.[5] If a tennis ball were one byte of information, a zettabyte-sized ball would be around the size of one earth. IDC recommends that businesses and organizations move now to create policies, tools, and standards to accommodate the approaching tidal wave of digital data and information.[6]

DATA MANAGEMENT

Without data and the ability to process it, an organization could not successfully complete most business activities. It could not pay employees, send out bills, order new inventory, or produce information to assist managers in decision making. Recall that data consists of raw facts, such as employee numbers and sales figures. For data to be transformed into useful information, it must first be organized in a meaningful way.

The Hierarchy of Data

Data is generally organized in a hierarchy that begins with the smallest piece of data used by computers (a bit) and progresses through the hierarchy to a database. A bit (a binary digit) represents a circuit that is either on or off. Bits can be organized into units called *bytes*. A byte is typically eight bits. Each byte represents a **character**, which is the basic building block of information. A character can be an uppercase letter (A, B, C… Z), lowercase letter (a, b, c… z), numeric digit (0, 1, 2… 9), or special symbol (., !, [+], [-], /, …).

character
A basic building block of information, consisting of uppercase letters, lowercase letters, numeric digits, or special symbols.

Characters can be combined to form a field. A **field** is typically a name, number, or combination of characters that describes an aspect of a business object (such as an employee, a location, or a truck) or activity (such as a sale). In addition to being entered into a database, fields can be computed from other fields. *Computed fields* include the total, average, maximum, and minimum values. A collection of related data fields is a **record**. By combining descriptions of the characteristics of an object or activity, a record can provide a complete description of the object or activity. For instance, an employee record is a collection of fields about one employee. One field includes the employee's name, another field contains the address, and still others the phone number, pay rate, earnings made to date, and so forth. A collection of related records is a **file**—for example, an employee file is a collection of all company employee records. Likewise, an inventory file is a collection of all inventory records for a particular company or organization. Some database software refers to files as tables.

field
Typically a name, number, or combination of characters that describes an aspect of a business object or activity.

record
A collection of related data fields.

file
A collection of related records.

At the highest level of this hierarchy is a *database*, a collection of integrated and related files. Together, bits, characters, fields, records, files, and databases form the **hierarchy of data** (see Figure 5.1). Characters are combined to make a field, fields are combined to make a record, records are combined to make a file, and files are combined to make a database. A database houses not only all these levels of data but also the relationships among them.

hierarchy of data
Bits, characters, fields, records, files, and databases.

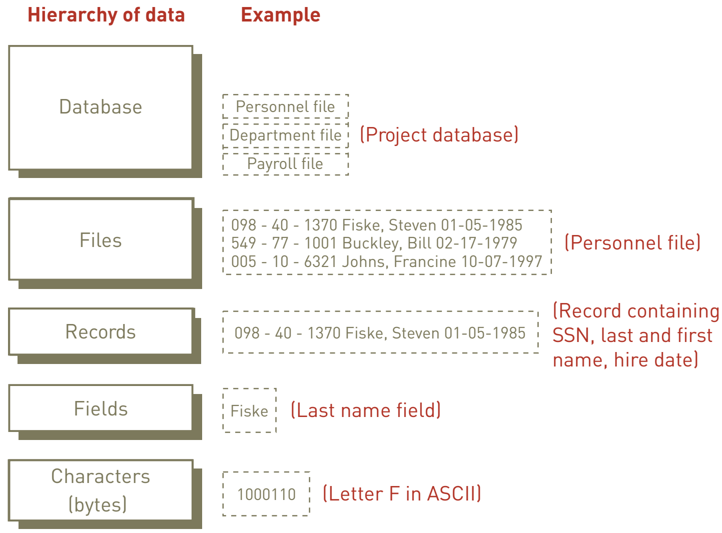
Data Entities, Attributes, and Keys

Entities, attributes, and keys are important database concepts. An **entity** is a generalized class of people, places, or things (objects) for which data is collected, stored, and maintained. Examples of entities include employees, inventory, and customers. Most organizations organize and store data as entities.

entity
A generalized class of people, places, or things for which data is collected, stored, and maintained.

Figure 5.1

The Hierarchy of Data



attribute
A characteristic of an entity.

data item
The specific value of an attribute.

An **attribute** is a characteristic of an entity. For example, employee number, last name, first name, hire date, and department number are attributes for an employee (see Figure 5.2). The inventory number, description, number of units on hand, and location of the inventory item in the warehouse are attributes for items in inventory. Customer number, name, address, phone number, credit rating, and contact person are attributes for customers. Attributes are usually selected to reflect the relevant characteristics of entities such as employees or customers. The specific value of an attribute, called a **data item**, can be found in the fields of the record describing an entity.

Figure 5.2

Keys and Attributes

The key field is the employee number. The attributes include last name, first name, hire date, and department number.

Employee #	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

ENTITIES (records)

KEY FIELD

ATTRIBUTES (fields)

Most organizations use attributes and data items. Many governments use attributes and data items to help in criminal investigations. The United States Federal Bureau of Investigation is building the "world's largest computer database of peoples' physical characteristics."[7] At a cost of $1 billion, the database management system named Next Generation Identification will catalog digital images of faces, fingerprints, and palm prints of U.S. citizens and visitors. Each person in the database is an entity, each biometric category is an attribute, and each image is a data item. The information will be used as a forensics tool and to increase homeland security.

key
A field or set of fields in a record that is used to identify the record.

primary key
A field or set of fields that uniquely identifies the record.

As discussed, a collection of fields about a specific object is a record. A **key** is a field or set of fields in a record that identifies the record. A **primary key** is a field or set of fields that uniquely identifies the record. No other record can have the same primary key. The primary key is used to distinguish records so that they can be accessed, organized, and manipulated. For an employee record, such as the one shown in Figure 5.2, the employee number is an example of a primary key.

Locating a particular record that meets a specific set of criteria might be easier and faster using a combination of secondary keys. For example, a customer might call a mail-order company to place an order for clothes. If the customer does not know the correct primary key (such as a customer number), a secondary key (such as last name) can be used. In this case, the order clerk enters the last name, such as Adams. If several customers have a last name of Adams, the clerk can check other fields, such as address, first name, and so on, to find the correct customer record. After locating the correct customer record, the order can be completed and the clothing items shipped to the customer.

The Database Approach

At one time, applications used specific files. For example, a payroll application would use a payroll file. In other words, each application used files dedicated to that application. This approach to data management, whereby separate data files are created and stored for each application program, is called the **traditional approach to data management**.

Today, most organizations use the **database approach to data management**, where multiple application programs share a pool of related data. A database offers the ability to share data and information resources. Federal databases, for example, often include the results of DNA tests as an attribute for convicted criminals. The information can be shared with law enforcement officials around the country.

To use the database approach to data management, additional software—a database management system (DBMS)—is required. As previously discussed, a DBMS consists of a group of programs that can be used as an interface between a database and the user of the database and application programs. Typically, this software acts as a buffer between the application programs and the database itself. Figure 5.3 illustrates the database approach.

traditional approach to data management
An approach whereby separate data files are created and stored for each application program.

database approach to data management
An approach whereby a pool of related data is shared by multiple application programs.

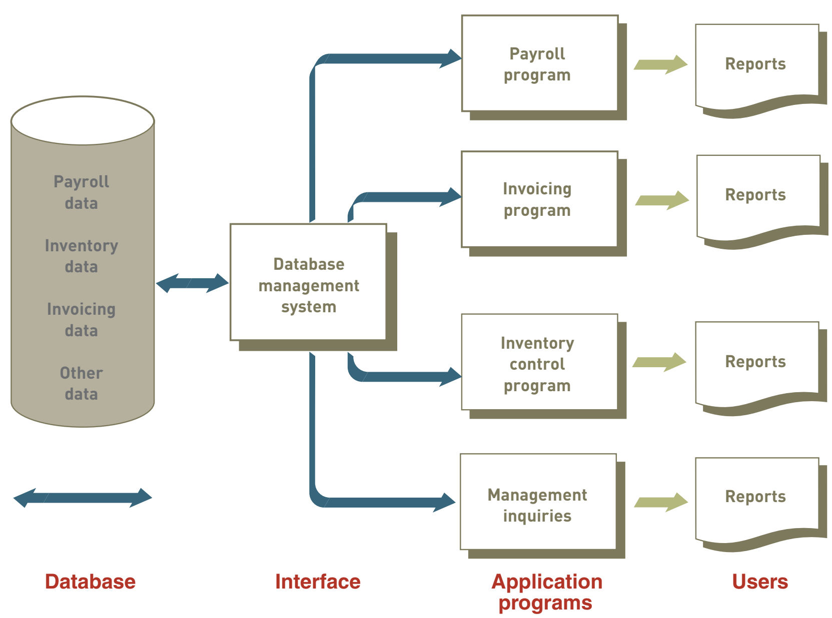


Figure 5.3
The Database Approach to Data Management

Table 5.1 lists some of the primary advantages of the database approach, and Table 5.2 lists some disadvantages.

Advantages	Explanation
Improved strategic use of corporate data	Accurate, complete, up-to-date data can be made available to decision makers where, when, and in the form they need it. The database approach can also give greater visibility to the organization's data resource.
Reduced data redundancy	Data is organized by the DBMS and stored in only one location. This results in more efficient use of system storage space.
Improved data integrity	With the traditional approach, some changes to data were not reflected in all copies of the data kept in separate files. The database approach prevents this problem because no separate files contain copies of the same piece of data.
Easier modification and updating	The DBMS coordinates data modifications and updates. Programmers and users do not have to know where the data is physically stored. Data is stored and modified once. Modification and updating is also easier because the data is commonly stored in only one location.
Data and program independence	The DBMS organizes the data independently of the application program, so the application program is not affected by the location or type of data. Introduction of new data types not relevant to a particular application does not require rewriting that application to maintain compatibility with the data file.
Better access to data and information	Most DBMSs have software that makes it easy to access and retrieve data from a database. In most cases, users give simple commands to get important information. Relationships between records can be more easily investigated and exploited, and applications can be more easily combined.
Standardization of data access	A standardized, uniform approach to database access means that all application programs use the same overall procedures to retrieve data and information.
A framework for program development	Standardized database access procedures can mean more standardization of program development. Because programs go through the DBMS to gain access to data in the database, standardized database access can provide a consistent framework for program development. In addition, each application program need address only the DBMS, not the actual data files, reducing application development time.
Better overall protection of the data	Accessing and using centrally located data is easier to monitor and control. Security codes and passwords can ensure that only authorized people have access to particular data and information in the database, thus ensuring privacy.
Shared data and information resources	The cost of hardware, software, and personnel can be spread over many applications and users. This is a primary feature of a DBMS.

Table 5.1

Advantages of the Database Approach

Table 5.2

Disadvantages of the Database Approach

Disadvantages	Explanation
More complexity	DBMSs can be difficult to set up and operate. Many decisions must be made correctly for the DBMS to work effectively. In addition, users have to learn new procedures to take full advantage of a DBMS.
More difficult to recover from a failure	With the traditional approach to file management, a failure of a file affects only a single program. With a DBMS, a failure can shut down the entire database.
More expensive	DBMSs can be more expensive to purchase and operate. The expense includes the cost of the database and specialized personnel, such as a database administrator, who is needed to design and operate the database. Additional hardware might also be required.

Many modern databases serve entire enterprises, encompassing much of the data of the organization. Often, distinct yet related databases are linked to provide enterprise-wide databases. For example, many Wal-Mart stores include in-store medical clinics for customers. Wal-Mart uses a centralized electronic health records database that stores the information of all patients across all stores.[8] The database is interconnected with the main Wal-Mart database to provide information about customers' interactions with the clinics and stores. The Ethical and Societal Issues box provides more information about databases used for electronic health record systems.

ETHICAL AND SOCIETAL ISSUES

Web-Based Electronic Health Record Systems

The United States federal government is pushing for most Americans to have their medical records stored in electronic form by 2014. Electronic health record (EHR) systems store patient records in a central database that can be accessed by many physicians at more than one location. Such a system eliminates problems caused by duplicate records at different physician offices, avoids having to fill out a new patient history with each new physician visited by the patient, and reduces errors made by incorrectly deciphering handwritten notes and prescriptions. Electronic records can make for a better and healthier world. However, the cost of moving to electronic systems is prohibitive, especially for small medical practices. At this point, only ten percent of small medical offices and five percent of solo practitioners have moved to EHR systems.

Although the government is introducing financial incentives to encourage physicians to use EHR systems, some big companies that aren't typically associated with healthcare are becoming involved—particularly Microsoft and Google. Approximately 52 percent of adults look to the Web when seeking health advice. Google and Microsoft believe that they can better assist health consumers by providing them with a robust tool for managing their health records. Microsoft's tool is named HealthVault, while Google's is named Google Health. The companies see their EHR systems as a solution to the government's problem for finding a low-cost records system designed for both physicians and patients.

John D. Halamka, a doctor and CIO of the Harvard Medical School, thinks systems in which the patient manages the information, such as those proposed by Microsoft and Google, are the inevitable future. "Patients will ultimately be the stewards of their own information," Halamka stated. "In the future, healthcare will be a much more collaborative process between patients and doctors."

Google agrees that patients should be in charge. A statement at Google Health's welcome page reads, "At Google, we feel patients should be in charge of their health information, and they should be able to grant their healthcare providers, family members, or whomever they choose, access to this information. Google Health was developed to meet this need."

But just how private and secure will our medical records be when stored in Web-accessible databases, protected only by one password? Privacy and security concerns are raised both by corporate access to private records by Microsoft and Google and outsider access by hackers. It is likely that both companies will use automated systems to target advertising at individuals based on medical records, just as Google's Gmail places ads next to e-mail messages based on the message contents. Unauthorized users might also be able to access records stored on a network that billions of users around the world use.

Another problem that complicates Google and Microsoft's involvement is that third-party medical record services are not covered by the Health Insurance Portability and Accountability Act (HIPAA). HIPAA provides strict standards for keeping medical records private. If a patient chooses to use Microsoft or Google to store medical records, those records would no longer be protected by the standards imposed by HIPAA in its current form.

As in similar cases, patients should weigh the costs in terms of privacy and security against the benefits of convenience and data reliability. Meanwhile, the software vendors need to work to build higher levels of security, privacy assurances, and customer trust.

Discussion Questions

1. Why does the U.S. federal government want to move health records to electronic systems?
2. What benefits and risks are offered by Web-based health records management systems like Google Health?

Critical Thinking Questions

1. How might Google and Microsoft reassure users about the privacy and security issues posed in this sidebar?
2. Would you consider registering for Google Health? Why or why not?

Sources: Lohr, Steve, "Google and Microsoft Look to Change Health Care," *New York Times*, August 14, 2007, *www.nytimes.com/2007/08/14/technology/14healthnet.html*, AP Staff, "Google ventures into health records biz," *CNN.com*, February 21, 2008, *www.cnn.com/2008/TECH/02/21/google.records.ap*.

DATA MODELING AND DATABASE CHARACTERISTICS

Because today's businesses have so many elements, they must keep data organized so that it can be used effectively. A database should be designed to store all data relevant to the business and provide quick access and easy modification. Moreover, it must reflect the business processes of the organization. When building a database, an organization must carefully consider these questions:

- **Content.** What data should be collected and at what cost?
- **Access.** What data should be provided to which users and when?
- **Logical structure.** How should data be arranged so that it makes sense to a given user?
- **Physical organization.** Where should data be physically located?

Data Modeling

Key considerations in organizing data in a database include determining what data to collect in the database, who will have access to it, and how they might want to use the data. After determining these details, an organization can create a database. Building a database requires two different types of designs: a logical design and a physical design. The *logical design* of a database is an abstract model of how the data should be structured and arranged to meet an organization's information needs. The logical design involves identifying relationships among the data items and grouping them in an orderly fashion. Because databases provide both input and output for information systems throughout a business, users from all functional areas should assist in creating the logical design to ensure that their needs are identified and addressed. *Physical design* starts from the logical database design and fine-tunes it for performance and cost considerations (such as improved response time, reduced storage space, and lower operating cost). For example, the database administrator at Intermountain Healthcare in Salt Lake City, Utah, combined the databases of 21 hospitals and 100 clinics into one integrated system, saving the organization the cost of dozens of servers, and providing new and improved services.[9] The person who fine-tunes the physical design must have an in-depth knowledge of the DBMS. For example, the logical database design might need to be altered so that certain data entities are combined, summary totals are carried in the data records rather than calculated from elemental data, and some data attributes are repeated in more than one data entity. These are examples of **planned data redundancy**, which improves the system performance so that user reports or queries can be created more quickly.

planned data redundancy
A way of organizing data in which the logical database design is altered so that certain data entities are combined, summary totals are carried in the data records rather than calculated from elemental data, and some data attributes are repeated in more than one data entity to improve database performance.

One of the tools database designers use to show the logical relationships among data is a data model. A **data model** is a diagram of entities and their relationships. Data modeling usually involves understanding a specific business problem and analyzing the data and information needed to deliver a solution. When done at the level of the entire organization, this is called enterprise data modeling. **Enterprise data modeling** is an approach that starts by investigating the general data and information needs of the organization at the strategic level, and then examines more specific data and information needs for the various functional areas and departments within the organization. Various models have been developed to help managers and database designers analyze data and information needs. An entity-relationship diagram is an example of such a data model.

data model
A diagram of data entities and their relationships.

enterprise data modeling
Data modeling done at the level of the entire enterprise.

Entity-relationship (ER) diagrams use basic graphical symbols to show the organization of and relationships between data. In most cases, boxes in ER diagrams indicate data items or entities contained in data tables, and diamonds show relationships between data items and entities. In other words, ER diagrams show data items in tables (entities) and the ways they are related.

entity-relationship (ER) diagrams
Data models that use basic graphical symbols to show the organization of and relationships between data.

ER diagrams help ensure that the relationships among the data entities in a database are correctly structured so that any application programs developed are consistent with business operations and user needs. In addition, ER diagrams can serve as reference documents after a database is in use. If changes are made to the database, ER diagrams help design them. Figure 5.4 shows an ER diagram for an order database. In this database design, one salesperson serves many customers. This is an example of a one-to-many relationship, as indicated by the one-to-many symbol (the "crow's-foot") shown in Figure 5.4. The ER diagram also shows that each customer can place one-to-many orders, each order includes one-to-many line items, and many line items can specify the same product (a many-to-one relationship). This database can also have one-to-one relationships. For example, one order generates one invoice.

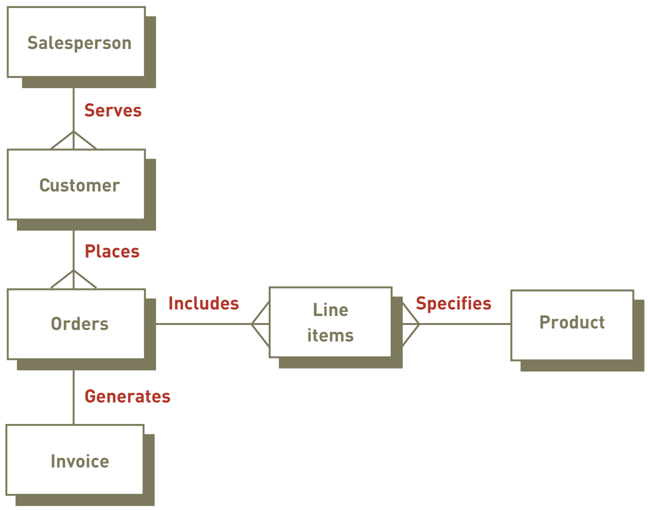


Figure 5.4

An Entity-Relationship (ER) Diagram for a Customer Order Database

Development of ER diagrams helps ensure that the logical structure of application programs is consistent with the data relationships in the database.

The Relational Database Model

Although there are a number of different database models, including flat files, hierarchical, and network models, the **relational model** has become the most popular, and use of this model will continue to increase. The relational model describes data using a standard tabular format. In a database structured according to the relational model, all data elements are placed in two-dimensional tables, called *relations*, which are the logical equivalent of files. The tables in relational databases organize data in rows and columns, simplifying data access and manipulation. It is normally easier for managers to understand the relational model (see Figure 5.5) than other database models.

relational model
A database model that describes data in which all data elements are placed in two-dimensional tables, called *relations*, which are the logical equivalent of files.

Databases based on the relational model include IBM DB2, Oracle, Sybase, Microsoft SQL Server, Microsoft Access, and MySQL. Oracle is currently the market leader in general-purpose databases, with over 40 percent of the $16.5 billion database market. IBM comes in second with about 21 percent, and Microsoft third with about 19 percent.[10]

In the relational model, each row (or record) of a table represents a data entity, with the columns (or fields) of the table representing attributes. Each attribute can accept only certain values. The allowable values for these attributes are called the **domain**. The domain for a particular attribute indicates what values can be placed in each column of the relational table. For instance, the domain for an attribute such as gender would be limited to male or female. A domain for pay rate would not include negative numbers. In this way, defining a domain can increase data accuracy.

domain
The allowable values for data attributes.

Manipulating Data

After entering data into a relational database, users can make inquiries and analyze the data. Basic data manipulations include selecting, projecting, and joining. **Selecting** involves eliminating rows according to certain criteria. Suppose a project table contains the project number, description, and department number for all projects a company is performing. The president of the company might want to find the department number for Project 226, a sales manual project. Using selection, the president can eliminate all rows but the one for Project 226 and see that the department number for the department completing the sales manual project is 598.

selecting
Manipulating data to eliminate rows according to certain criteria.

Figure 5.5

A Relational Database Model

In the relational model, all data elements are placed in two-dimensional tables, or relations. As long as they share at least one common element, these relations can be linked to output useful information. Note that some organizations might use employee number instead of Social Security number (SSN) in Data Tables 2 and 3.

Data Table 1: Project Table

Project	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales manual	598

Data Table 2: Department Table

Dept.	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data Table 3: Manager Table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

projecting
Manipulating data to eliminate columns in a table.

joining
Manipulating data to combine two or more tables.

linking
Data manipulation that combines two or more tables using common data attributes to form a new table with only the unique data attributes.

Projecting involves eliminating columns in a table. For example, a department table might contain the department number, department name, and Social Security number (SSN) of the manager in charge of the project. A sales manager might want to create a new table with only the department number and the Social Security number of the manager in charge of the sales manual project. The sales manager can use projection to eliminate the department name column and create a new table containing only department number and SSN.

Joining involves combining two or more tables. For example, you can combine the project table and the department table to create a new table with the project number, project description, department number, department name, and Social Security number for the manager in charge of the project.

As long as the tables share at least one common data attribute, the tables in a relational database can be **linked** to provide useful information and reports. Being able to link tables to each other through common data attributes is one of the keys to the flexibility and power of relational databases. Suppose the president of a company wants to find out the name of the manager of the sales manual project and the length of time the manager has been with the company. Assume that the company has the manager, department, and project tables shown in Figure 5.5. A simplified ER diagram showing the relationship between these tables is shown in Figure 5.6. Note the crow's-foot by the project table. This indicates that a department can have many projects. The president would make the inquiry to the database, perhaps via a personal computer. The DBMS would start with the project description and search the project table to find out the project's department number. It would then use the department number to search the department table for the manager's Social Security number. The department number is also in the department table and is the common element that links the project table to the department table. The DBMS uses the manager's Social Security number to search the manager table for the manager's hire date. The manager's Social Security number is the common element between the department table and the manager table. The final result is that the manager's name and hire date are presented to the president as a response to the inquiry (see Figure 5.7).

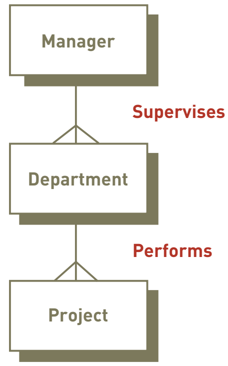
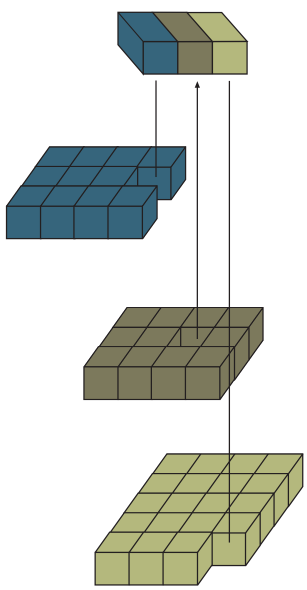


Figure 5.6

A Simplified ER Diagram Showing the Relationship Between the Manager, Department, and Project Tables



Data Table 1: Project Table

Project number	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales manual	598

Data Table 2: Department Table

Dept. number	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data Table 3: Manager Table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

Figure 5.7

Linking Data Tables to Answer an Inquiry

In finding the name and hire date of the manager working on the sales manual project, the president needs three tables: project, department, and manager. The project description (Sales manual) leads to the department number (598) in the project table, which leads to the manager's SSN (098-40-1370) in the department table, which leads to the manager's name (Fiske) and hire date (01-05-1985) in the manager table. Again, note that some organizations might use employee number instead of Social Security number (SSN).

One of the primary advantages of a relational database is that it allows tables to be linked, as shown in Figure 5.7. This linkage is especially useful when information is needed from multiple tables. For example, the manager's Social Security number is maintained in the manager table. If the Social Security number is needed, it can be obtained by linking to the manager table.

The relational database model is by far the most widely used. It is easier to control, more flexible, and more intuitive than other approaches because it organizes data in tables. As shown in Figure 5.8, a relational database management system, such as Access, provides tips and tools for building and using database tables. In this figure, the database displays information about data types and indicates that additional help is available. The ability to link relational tables also allows users to relate data in new ways without having to redefine complex relationships. Because of the advantages of the relational model, many companies use it for large corporate databases, such as those for marketing and accounting. The relational model can also be used with personal computers and mainframe systems. A travel reservation company, for example, can develop a fare-pricing system by using relational database technology that can handle millions of daily queries from online travel companies, such as Expedia, Travelocity, and Orbitz.

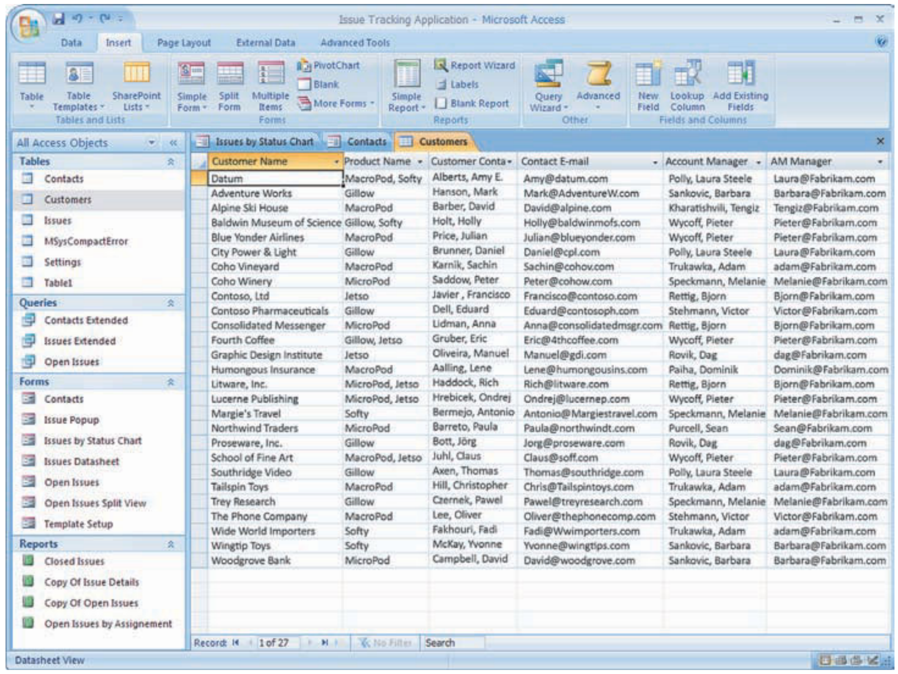


Figure 5.8

Building and Modifying a Relational Database

Relational databases provide many tools, tips, and shortcuts to simplify the process of creating and modifying a database.

(Source: Courtesy of Microsoft Corporation.)

Data Cleanup

data cleanup
The process of looking for and fixing inconsistencies to ensure that data is accurate and complete.

As discussed in Chapter 1, valuable data is accurate, complete, economical, flexible, reliable, relevant, simple, timely, verifiable, accessible, and secure. The database must also be properly designed. The purpose of **data cleanup** is to develop data with these characteristics. Consider a database for a fitness center designed to track member dues. The table contains the attribute name, phone number, gender, dues paid, and date paid (see Table 5.3). As the records in Table 5.3 show, Anita Brown and Sim Thomas have paid their dues in September. Sim has paid his dues in two installments. Note that no primary key uniquely identifies each record. As you will see next, this problem must be corrected.

Table 5.3

Fitness Center Dues

Name	Phone	Gender	Dues Paid	Date Paid
Brown, A.	468-3342	Female	$30	September 15
Thomas, S.	468-8788	Male	$15	September 15
Thomas, S.	468-5238	Male	$15	September 25

Because Sim Thomas has paid dues twice in September, the data in the database is now redundant. The name, phone number, and gender for Thomas are repeated in two records. Notice that the data in the database is also inconsistent: Thomas has changed his phone number, but only one of the records reflects this change. Further reducing this database's reliability is the lack of a primary key to uniquely identify Sim Thomas's record. The first Thomas could be Sim Thomas, but the second might be Steve Thomas. These problems and irregularities in data are called *anomalies*. Data anomalies often result in incorrect information, causing database users to be misinformed about actual conditions. Anomalies must be corrected.

To solve these problems in the fitness center's database, we can add a primary key, such as member number, and put the data into two tables: a Fitness Center Members table with gender, phone number, and related information, and a Dues Paid table with dues paid and date paid (see Tables 5.4 and 5.5). Both tables include the member number attribute so that they can be linked.

Table 5.4

Fitness Center Members

Member No.	Name	Phone	Gender
SN123	Brown, A.	468-3342	Female
SN656	Thomas, S.	468-5238	Male

Table 5.5

Dues Paid

Member No.	Dues Paid	Date Paid
SN123	$30	September 15
SN656	$15	September 15
SN656	$15	September 25

The relations in Table 5.4 and Table 5.5 reduce the redundancy and eliminate the potential problem of having two different phone numbers for the same member. Also note that the member number gives each record in the Fitness Center Members table a primary key. Because the Dues Paid table lists two payment entries ($15 each) with the same member number (SN656), one person clearly made the payments, not two different people. Formalized approaches, such as *database normalization*, are often used to clean up problems with data.

DATABASE MANAGEMENT SYSTEMS

Creating and implementing the right database system ensures that the database will support both business activities and goals. But how do we actually create, implement, use, and update a database? The answer is found in the database management system. As discussed earlier, a DBMS is a group of programs used as an interface between a database and application programs or a database and the user. The capabilities and types of database systems, however, vary considerably. For example, visitors to the Baseball Hall of Fame in Cooperstown, New York, use a DBMS to search baseball highlight films from famous games and plays.[11] DBMSs are used to manage all kinds of data for all kinds of purposes.

Overview of Database Types

Database management systems can range from small, inexpensive software packages to sophisticated systems costing hundreds of thousands of dollars. The following sections discuss a few popular alternatives. See Figure 5.9 for one example.

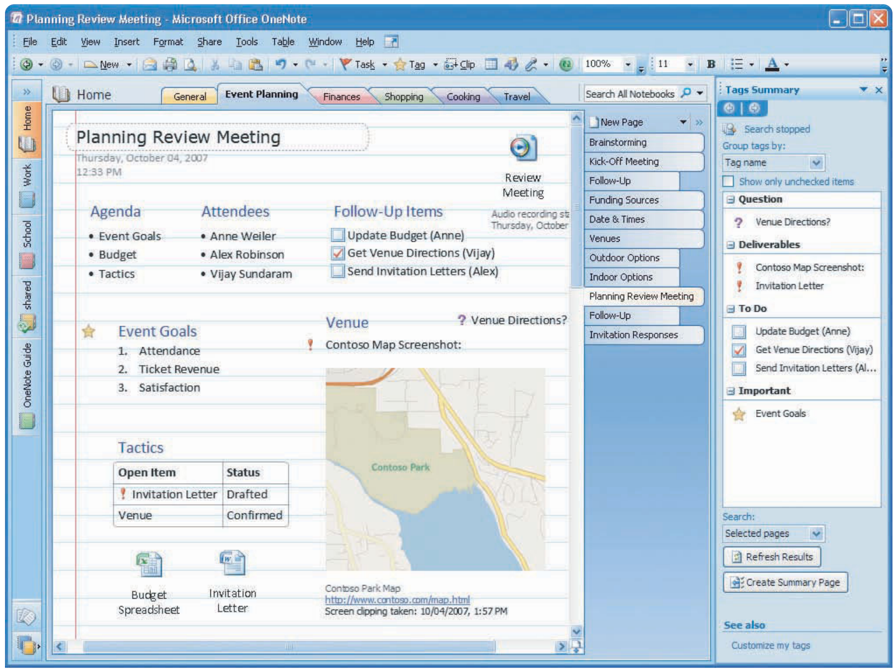
Flat File

A flat file is a simple database program whose records have no relationship to one another. Flat file databases are often used to store and manipulate a single table or file, and do not use any of the database models discussed previously, such as the relational model. Many spreadsheet and word processing programs have flat file capabilities. These software packages can sort tables and make simple calculations and comparisons. Microsoft OneNote is designed to let people put ideas, thoughts, and notes into a computer file. In OneNote, each note can be placed anywhere on a page or in a box on a page, called a *container*. Pages are organized into sections and subsections that appear as colored tabs. After you enter a note, you can retrieve, copy, and paste it into other applications, such as word processing and spreadsheet programs. Microsoft uses OneNote as the primary technology for its management training classes. OneNote allows managers-in-training to collect photos, handwritten notes, online content, and audio recordings in one flat file.[12] OneNote enables Microsoft to offer training to a larger number of managers, while saving $360,000 per year in printed training materials.

Figure 5.9

Microsoft OneNote

Microsoft OneNote lets you gather any type of information and then retrieve, copy, and paste the information into other applications, such as word processing and spreadsheet programs.



Similar to OneNote, Evernote is a free database that can store notes and other pieces of information. Considering the amount of information today's high-capacity hard disks can store, the popularity of databases that can handle unstructured data will continue to grow.

Single User

A database installed on a personal computer is typically meant for a single user. Microsoft Office Access and FileMaker Pro are designed to support single-user implementations. Microsoft InfoPath is another example of a database program that supports a single user. This software is part of the Microsoft Office suite, and it helps people collect and organize information from a variety of sources. InfoPath has built-in forms that can be used to enter expense information, timesheet data, and a variety of other information.

Multiple Users

Small, midsize, and large businesses need multiuser DBMSs to share information throughout the organization over a network. These more powerful, expensive systems allow dozens or hundreds of people to access the same database system at the same time. Popular vendors for multiuser database systems include Oracle, Microsoft, Sybase, and IBM. Many single-user databases, such as Microsoft Access, can be implemented for multiuser support over a network, though they often are limited in the amount of users they can support.

All DBMSs share some common functions, such as providing a user view, physically storing and retrieving data in a database, allowing for database modification, manipulating data, and generating reports. These DBMSs can handle the most complex data-processing tasks, and because they are accessed over a network, one database can serve many locations around the world. For example, Surya Roshni Ltd is a major manufacturer of lighting products based in New Delhi, India, with a global reach. One Oracle database stored on servers in New Delhi provides corporate information to associates around the world.[13]

Providing a User View

Because the DBMS is responsible for access to a database, one of the first steps in installing and using a large database involves telling the DBMS the logical and physical structure of the data and relationships among the data in the database for each user. This description is called a **schema** (as in schematic diagram). Large database systems, such as Oracle, typically

schema
A description of the entire database.

use schemas to define the tables and other database features associated with a person or user. A schema can be part of the database or a separate schema file. The DBMS can reference a schema to find where to access the requested data in relation to another piece of data.

Creating and Modifying the Database

Schemas are entered into the DBMS (usually by database personnel) via a data definition language. A **data definition language (DDL)** is a collection of instructions and commands used to define and describe data and relationships in a specific database. A DDL allows the database's creator to describe the data and relationships that are to be contained in the schema. In general, a DDL describes logical access paths and logical records in the database. Figure 5.10 shows a simplified example of a DDL used to develop a general schema. The *Xs* in Figure 5.10 reveal where specific information concerning the database should be entered. File description, area description, record description, and set description are terms the DDL defines and uses in this example. Other terms and commands can be used, depending on the particular DBMS employed.

data definition language (DDL)
A collection of instructions and commands used to define and describe data and relationships in a specific database.

```
SCHEMA DESCRIPTION
SCHEMA NAME IS XXXX
AUTHOR      XXXX
DATE        XXXX
FILE DESCRIPTION
    FILE NAME IS XXXX
      ASSIGN XXXX
    FILE NAME IS XXXX
      ASSIGN XXXX
AREA DESCRIPTION
    AREA NAME IS XXXX
RECORD DESCRIPTION
    RECORD NAME IS XXXX
    RECORD ID IS XXXX
    LOCATION MODE IS XXXX
    WITHIN XXXX AREA FROM XXXX THRU XXXX
SET DESCRIPTION
    SET NAME IS XXXX
    ORDER IS XXXX
    MODE IS XXXX
    MEMBER IS XXXX
    .
    .
    .
```

Figure 5.10

Using a Data Definition Language to Define a Schema

Another important step in creating a database is to establish a **data dictionary**, a detailed description of all data used in the database. The data dictionary contains the following data:

- Name of the data item
- Aliases or other names that may be used to describe the item
- Range of values that can be used
- Type of data (such as alphanumeric or numeric)
- Amount of storage needed for the item
- Notation of the person responsible for updating it and the various users who can access it
- List of reports that use the data item

data dictionary
A detailed description of all the data used in the database.

A data dictionary can also include a description of data flows, the way records are organized, and the data-processing requirements. Figure 5.11 shows a typical data dictionary entry.

Figure 5.11

A Typical Data Dictionary Entry

NORTHWESTERN MANUFACTURING

PREPARED BY:	D. BORDWELL
DATE:	04 AUGUST 2007
APPROVED BY:	J. EDWARDS
DATE:	13 OCTOBER 2007
VERSION:	3.1
PAGE:	1 OF 1
DATA ELEMENT NAME:	PARTNO
DESCRIPTION:	INVENTORY PART NUMBER
OTHER NAMES:	PTNO
VALUE RANGE:	100 TO 5000
DATA TYPE:	NUMERIC
POSITIONS:	4 POSITIONS OR COLUMNS

For example, the information in a data dictionary for the part number of an inventory item can include the following data:

- Name of the person who made the data dictionary entry (D. Bordwell)
- Date the entry was made (August 4, 2007)
- Name of the person who approved the entry (J. Edwards)
- Approval date (October 13, 2007)
- Version number (3.1)
- Number of pages used for the entry (1)
- Part name (PARTNO)
- Part names that might be used (PTNO)
- Range of values (part numbers can range from 100 to 5,000)
- Type of data (numeric)
- Storage required (four positions are required for the part number)

A data dictionary is valuable in maintaining an efficient database that stores reliable information with no redundancy, and makes it easy to modify the database when necessary. Data dictionaries also help computer and system programmers who require a detailed description of data elements stored in a database to create the code to access the data.

Storing and Retrieving Data

One function of a DBMS is to be an interface between an application program and the database. When an application program needs data, it requests the data through the DBMS. Suppose that to calculate the total price of a new car, an auto dealer pricing program needs price data on the engine option—six cylinders instead of the standard four cylinders. The application program requests this data from the DBMS. In doing so, the application program follows a logical access path. Next, the DBMS, working with various system programs, accesses a storage device, such as disk drives, where the data is stored. When the DBMS goes to this storage device to retrieve the data, it follows a path to the physical location (physical access path) where the price of this option is stored. In the pricing example, the DBMS might go to a disk drive to retrieve the price data for six-cylinder engines. This relationship is shown in Figure 5.12.

This same process is used if a user wants to get information from the database. First, the user requests the data from the DBMS. For example, a user might give a command, such as LIST ALL OPTIONS FOR WHICH PRICE IS GREATER THAN 200 DOLLARS. This is the logical access path (LAP). Then, the DBMS might go to the options price section of a disk to get the information for the user. This is the physical access path (PAP).

Two or more people or programs attempting to access the same record in the same database at the same time can cause a problem. For example, an inventory control program might attempt to reduce the inventory level for a product by ten units because ten units were just shipped to a customer. At the same time, a purchasing program might attempt to increase

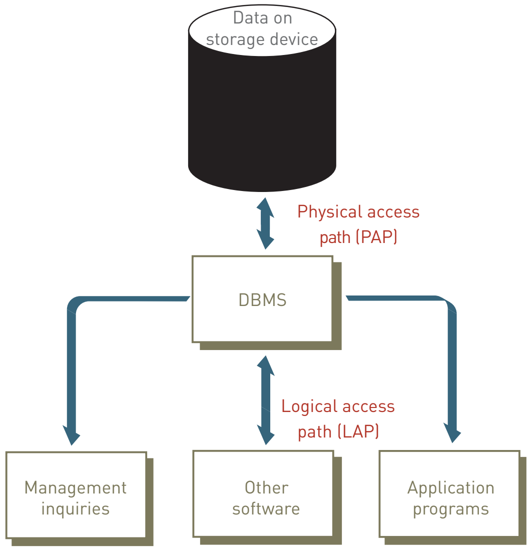


Figure 5.12

Logical and Physical Access Paths

the inventory level for the same product by 200 units because more inventory was just received. Without proper database control, one of the inventory updates might be incorrect, resulting in an inaccurate inventory level for the product. **Concurrency control** can be used to avoid this potential problem. One approach is to lock out all other application programs from access to a record if the record is being updated or used by another program.

concurrency control
A method of dealing with a situation in which two or more people need to access the same record in a database at the same time.

Manipulating Data and Generating Reports

After a DBMS has been installed, employees, managers, and consumers can use it to review reports and obtain important information. For example, the Food Allergen and Consumer Protection Act, effective in 2006, requires that food manufacturing companies generate reports on the ingredients, formulas, and food preparation techniques for the public. Using a DBMS, a company can easily manage this requirement.

Some databases use *Query-by-Example (QBE)*, which is a visual approach to developing database queries or requests. Like Windows and other GUI operating systems, you can perform queries and other database tasks by opening windows and clicking the data or features you want (see Figure 5.13).

In other cases, database commands can be used in a programming language. For example, C++ commands can be used in simple programs that will access or manipulate certain pieces of data in the database. Here's another example of a DBMS query: SELECT * FROM EMPLOYEE WHERE JOB_CLASSIFICATION = "C2". The * tells the program to include all columns from the EMPLOYEE table. In general, the commands that are used to manipulate the database are part of the **data manipulation language (DML)**. This specific language, provided with the DBMS, allows managers and other database users to access, modify, and make queries about data contained in the database to generate reports. Again, the application programs go through schemas and the DBMS before actually getting to the physically stored data on a device such as a disk.

data manipulation language (DML)
The commands that are used to manipulate the data in a database.

In the 1970s, D. D. Chamberlain and others at the IBM Research Laboratory in San Jose, California, developed a standardized data manipulation language called *Structured Query Language (SQL*, pronounced like *sequel*). The EMPLOYEE query shown earlier is written in SQL. In 1986, the American National Standards Institute (ANSI) adopted SQL as the standard query language for relational databases. Since ANSI's acceptance of SQL, interest in making SQL an integral part of relational databases on both mainframe and personal computers has increased. SQL has many built-in functions, such as average (AVG), the largest value (MAX), the smallest value (MIN), and others. Table 5.6 contains examples of SQL commands.

Figure 5.13

Query by Example

Some databases use Query-by-Example (QBE) to generate reports and information.

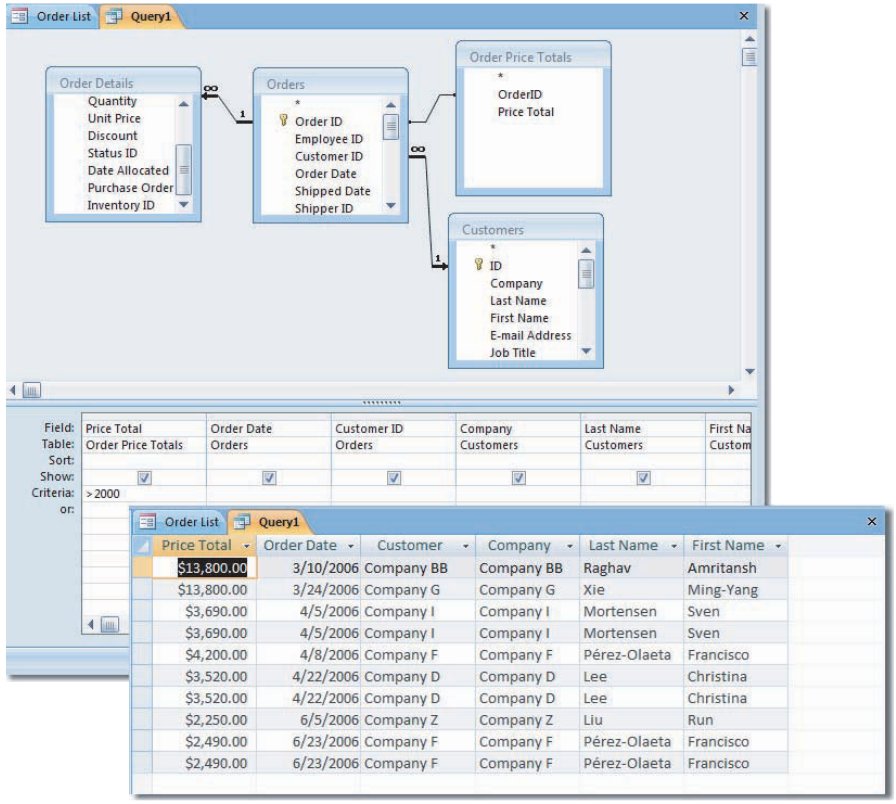


Table 5.6

Examples of SQL Commands

SQL Command	Description
SELECT ClientName, Debt FROM Client WHERE Debt > 1000	This query displays all clients (ClientName) and the amount they owe the company (Debt) from a database table called Client for clients who owe the company more than $1,000 (WHERE Debt > 1000).
SELECT ClientName, ClientNum, OrderNum FROM Client, Order WHERE Client.ClientNum=Order.ClientNum	This command is an example of a join command that combines data from two tables: the client table and the order table (FROM Client, Order). The command creates a new table with the client name, client number, and order number (SELECT ClientName, ClientNum, OrderNum). Both tables include the client number, which allows them to be joined. This is indicated in the WHERE clause, which states that the client number in the client table is the same as (equal to) the client number in the order table (WHERE Client.Client Num= Order.ClientNum).
GRANT INSERT ON Client to Guthrie	This command is an example of a security command. It allows Bob Guthrie to insert new values or rows into the Client table.

SQL lets programmers learn one powerful query language and use it on systems ranging from PCs to the largest mainframe computers (see Figure 5.14). Programmers and database users also find SQL valuable because SQL statements can be embedded into many programming languages, such as the widely used C++ and COBOL languages. Because SQL uses standardized and simplified procedures for retrieving, storing, and manipulating data in a database system, the popular database query language can be easy to understand and use.

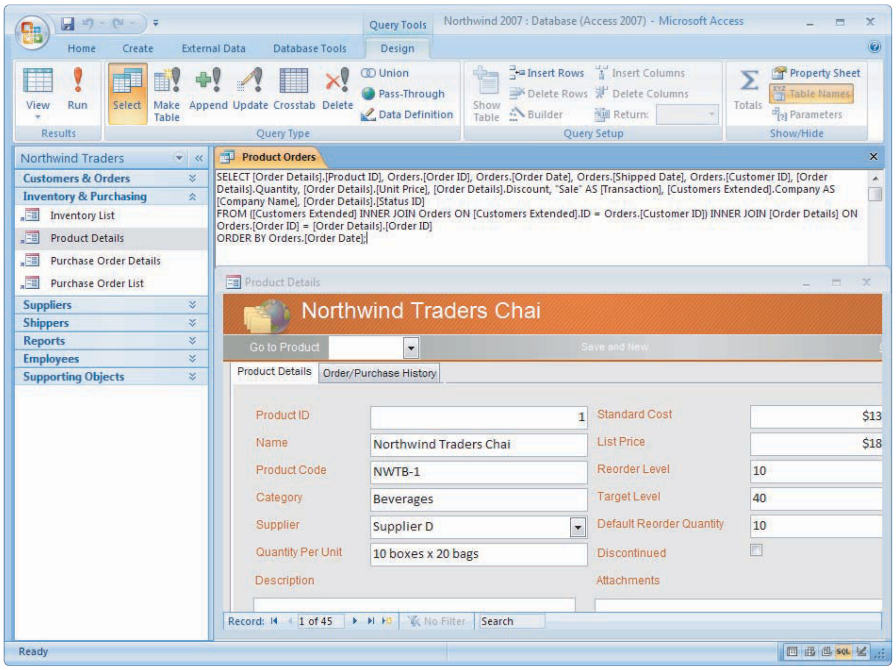


Figure 5.14

Structured Query Language

Structured Query Language (SQL) has become an integral part of most relational databases, as shown by this screen from Microsoft Access 2007.

After a database has been set up and loaded with data, it can produce desired reports, documents, and other outputs (see Figure 5.15). These outputs usually appear in screen displays or hard-copy printouts. The output-control features of a database program allow you to select the records and fields to appear in reports. You can also make calculations specifically for the report by manipulating database fields. Formatting controls and organization options (such as report headings) help you to customize reports and create flexible, convenient, and powerful information-handling tools.

Figure 5.15

Database Output

A database application offers sophisticated formatting and organization options to produce the right information in the right format.

A DBMS can produce a wide variety of documents, reports, and other output that can help organizations achieve their goals. The most common reports select and organize data to present summary information about some aspect of company operations. For example, accounting reports often summarize financial data such as current and past-due accounts. Many companies base their routine operating decisions on regular status reports that show the progress of specific orders toward completion and delivery.

Databases can also provide support to help executives and other people make better decisions. A database by Intellifit, for example, can be used to help shoppers make better decisions and get clothes that fit when shopping online. The database contains true sizes of apparel from various clothing companies that do business on the Web. The process starts when a customer's body is scanned into a database at one of the company's locations, typically in a shopping mall. About 200,000 measurements are taken to construct a 3-D image of the person's body shape. The database then compares the actual body dimensions with sizes given by Web-based clothing stores to get an excellent fit.[14]

Database Administration

Database systems require a skilled DBA. A DBA is expected to have a clear understanding of the fundamental business of the organization, be proficient in the use of selected database management systems, and stay abreast of emerging technologies and new design approaches. The role of the DBA is to plan, design, create, operate, secure, monitor, and maintain databases. Typically, a DBA has a degree in computer science or management information systems and some on-the-job training with a particular database product or more extensive experience with a range of database products. See Figure 5.16.

Figure 5.16

Database Administrator

The role of the database administrator (DBA) is to plan, design, create, operate, secure, monitor, and maintain databases.

(Source: BananaStock / Alamy.)

The DBA works with users to decide the content of the database—to determine exactly what entities are of interest and what attributes are to be recorded about those entities. Thus, personnel outside of IS must have some idea of what the DBA does and why this function is important. The DBA can play a crucial role in the development of effective information systems to benefit the organization, employees, and managers.

The DBA also works with programmers as they build applications to ensure that their programs comply with database management system standards and conventions. After the database is built and operating, the DBA monitors operations logs for security violations. Database performance is also monitored to ensure that the system's response time meets users' needs and that it operates efficiently. If there is a problem, the DBA attempts to correct it before it becomes serious.

data administrator
A nontechnical position responsible for defining and implementing consistent principles for a variety of data issues.

Some organizations have also created a position called the *data administrator*, a nontechnical, but important role that ensures that data is managed as an important organizational resource. The **data administrator** is responsible for defining and implementing consistent principles for a variety of data issues, including setting data standards and data definitions that apply across all the databases in an organization. For example, the data administrator would ensure that a term such as "customer" is defined and treated consistently in all corporate databases. This person also works with business managers to identify who should have read or update access to certain databases and to selected attributes within those databases. This information is then communicated to the database administrator for implementation. The data administrator can be a high-level position reporting to top-level managers.

Popular Database Management Systems

Some popular DBMSs for single users include Microsoft Access and FileMaker Pro. The complete database management software market encompasses software used by professional programmers that runs on midrange, mainframe, and supercomputers. The entire market, including IBM, Oracle, and Microsoft, generates billions of dollars per year in revenue. Although Microsoft rules in the desktop PC software market, its share of database software on larger computers is small.

Like other software products, a number of open-source database systems are available, including PostgreSQL and MySQL. Open-source software was described in Chapter 4. In addition, many traditional database programs are now available on open-source operating systems. The popular DB2 relational database from IBM, for example, is available on the Linux operating system. The Sybase IQ database and other databases are also available on the Linux operating system.

A new form of database system is emerging that some refer to as *Database as a Service* (*DaaS*) and others as Database 2.0. DaaS is similar to software as a service (SaaS). Recall that a SaaS system is one in which the software is stored on a service provider's servers and accessed by the client company over a network. In DaaS, the database is stored on a service provider's servers and accessed by the client over a network, typically the Internet. In DaaS, database administration is provided by the service provider. SaaS and DaaS are both part of the larger cloud computing trend. Recall from Chapter 3 that cloud computing uses a giant cluster of computers that serves as a host to run applications that require high-performance computing. In cloud computing, all information systems and data are maintained and managed by service providers and delivered over the Internet. Businesses and individuals are freed from having to install, service, maintain, upgrade, and safeguard their systems.

More than a dozen companies are moving in the DaaS direction. They include Google, Microsoft, Intuit, Serran Tech, MyOwnDB, and Trackvia.[15] XM Radio, Google, JetBlue Airways, Bank of America, Southwest Airlines, and others use QuickBase from service provider Intuit to manage their databases out of house.[16] JetBlue, for example, uses a DaaS from Intuit to organize and manage IT projects.[17] Because the database and DBMS are available from any Internet connection, those involved in managing and implementing systems development projects can record their progress and check on others' progress from any location.

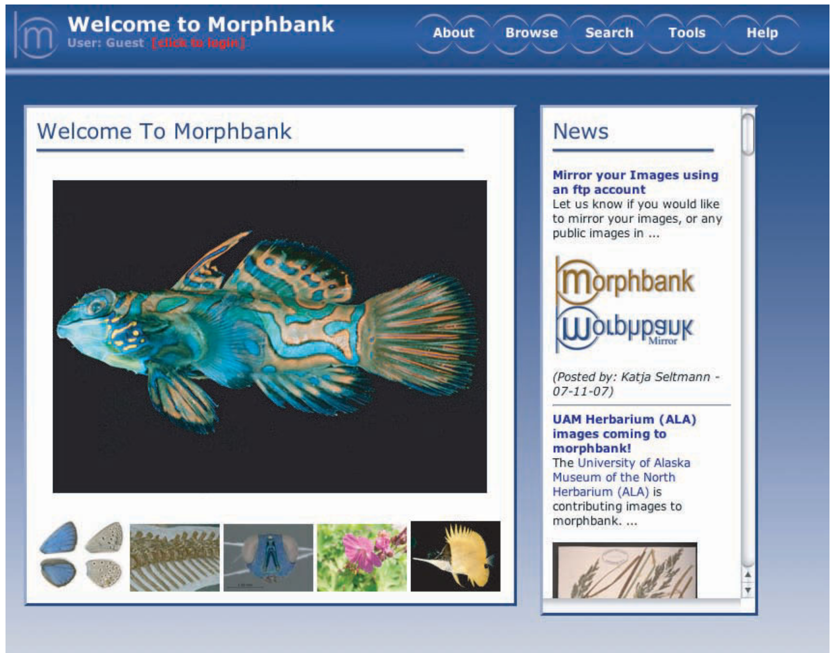
Special-Purpose Database Systems

In addition to the popular database management systems just discussed, some specialized database packages are used for specific purposes or in specific industries. For example, the Israeli Holocaust Database (*www.yadvashem.org*) is a special-purpose database available through the Internet and contains information on about three million people in 14 languages. A unique special-purpose DBMS for biologists called Morphbank (*www.morphbank.net*) allows researchers from around the world to continually update and expand a library of over 96,000 biological images to share with the scientific community and the public. The iTunes store music and video catalog is a special-purpose database system. When you search for your favorite artist, you are querying the database.

Selecting a Database Management System

The database administrator often selects the best database management system for an organization. The process begins by analyzing database needs and characteristics. The information needs of the organization affect the type of data that is collected and the type of database management system that is used. Important characteristics of databases include the following:

- **Database size.** The number of records or files in the database
- **Database cost.** The purchase or lease costs of the database
- **Concurrent users.** The number of people who need to use the database at the same time (the number of concurrent users)
- **Performance.** How fast the database is able to update records
- **Integration.** The ability to be integrated with other applications and databases
- **Vendor.** The reputation and financial stability of the database vendor



The Web-based Morphbank database allows scientists from around the world to upload and share biological and microscopic photographs and descriptions that support research in many areas.

(Source: *www.morphbank.net*)

For many organizations, database size doubles about every year or two. With the increasing use of digital media—images, video, and audio—data storage demands are growing exponentially. In fact, the volume of data being created has surpassed the world's available storage capacity.[18] The growing need for data storage has not escaped the notice of large technology companies such as Google and Microsoft, who are buying hundreds of acres of land and building huge data centers to support the world's data storage needs.[19] Meanwhile, many businesses and government agencies are working to consolidate data dispersed across the organization into smaller, more efficient centralized systems.

Using Databases with Other Software

Database management systems are often used with other software or the Internet. A DBMS can act as a front-end application or a back-end application. A *front-end application* is one that directly interacts with people or users. Marketing researchers often use a database as a front end to a statistical analysis program. The researchers enter the results of market questionnaires or surveys into a database. The data is then transferred to a statistical analysis program to determine the potential for a new product or the effectiveness of an advertising campaign. A *back-end application* interacts with other programs or applications; it only indirectly interacts with people or users. When people request information from a Web site, the Web site can interact with a database (the back end) that supplies the desired information. For example, you can connect to a university Web site to find out whether the university's library has a book you want to read. The Web site then interacts with a database that contains a catalog of library books and articles to determine whether the book you want is available.

DATABASE APPLICATIONS

Today's database applications manipulate the content of a database to produce useful information. Common manipulations are searching, filtering, synthesizing, and assimilating the data contained in a database, using a number of database applications. These applications allow users to link the company databases to the Internet, set up data warehouses and marts, use databases for strategic business intelligence, place data at different locations, use online processing and open connectivity standards for increased productivity, develop databases with the object-oriented approach, and search for and use unstructured data, such as graphics, audio, and video.

Linking the Company Database to the Internet

Linking databases to the Internet is one reason the Internet is so popular. A large percentage of corporate databases are accessed over the Internet through a standard Web browser. Being able to access bank account data, student transcripts, credit card bills, product catalogs, and a host of other data online is convenient for individual users, and increases effectiveness and efficiency for businesses and organizations. Amazon.com, Apple's iTunes store, eBay, and others have made billions of dollars by combining databases, the Internet, and smart business models.

As discussed in the Ethical and Societal Issues sidebar, Google is rolling out a DBMS that will provide patients and physicians with one storage location for all medical records, accessed through a Web browser.[20] Access to private medical information over the public Web has some privacy advocates concerned. However, the convenience that the system offers by dramatically reducing the number of paper forms to fill out and store, along with the reduction of clerical errors through streamlined data management procedures, has most in the field supporting the move to a centralized system. Google protects patient records with encryption and authentication technologies.

Developing a seamless integration of traditional databases with the Internet is often called a *semantic Web*. A semantic Web allows people to access and manipulate a number of traditional databases at the same time through the Internet. The World Wide Web Consortium has established standards for a semantic Web in hopes of some day evolving the Web into one big database that is easy to manage and traverse. Yahoo has recently announced its commitment to complying with the standards for a semantic Web.[21]

Although the semantic Web standards have not been embraced by all businesses, many software vendors—including IBM, Oracle, Microsoft, Macromedia, and Inline Internet Systems—are incorporating the Internet into their products. Such databases allow companies to create an Internet-accessible catalog, which is a database of items, descriptions, and prices. As evidenced by the Web, most companies are using these tools to take their business online.

In addition to the Internet, organizations are gaining access to databases through networks to find good prices and reliable service. Connecting databases to corporate Web sites and networks can lead to potential problems, however. A recent study found that nearly half a million database servers were vulnerable to attack over the Internet due to the lack of proper security measures.[22]

Data Warehouses, Data Marts, and Data Mining

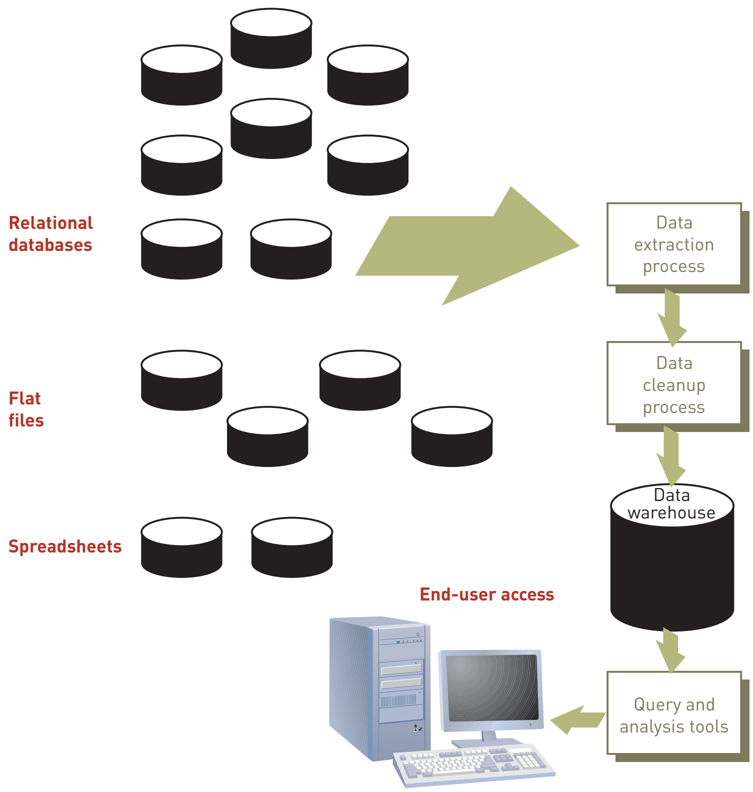
The raw data necessary to make sound business decisions is stored in a variety of locations and formats. This data is initially captured, stored, and managed by transaction processing systems that are designed to support the day-to-day operations of the organization. For decades, organizations have collected operational, sales, and financial data with their online transaction processing (OLTP) systems. The data can be used to support decision making using data warehouses, data marts, and data mining.

Data Warehouses

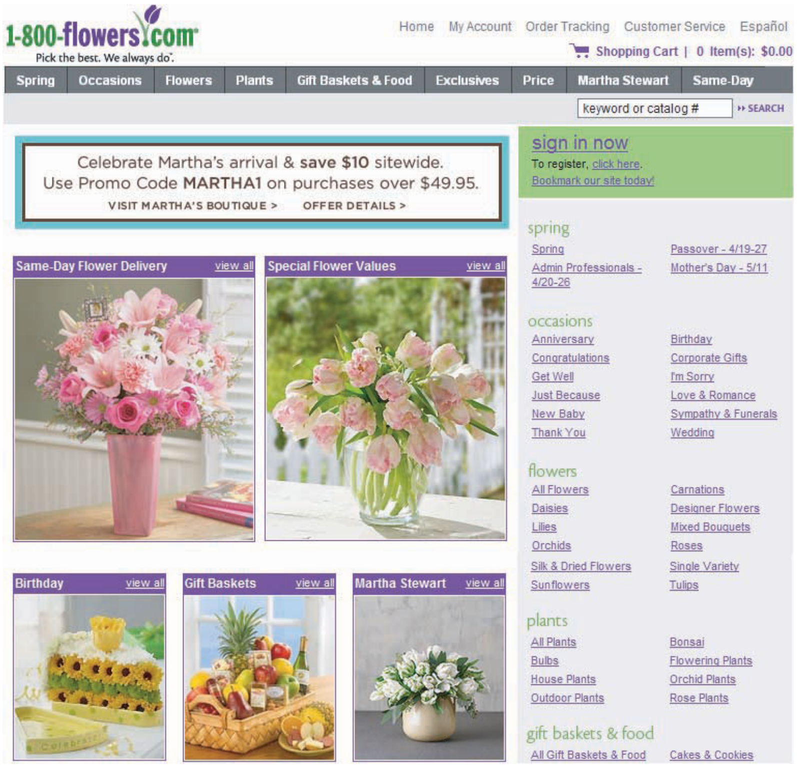
data warehouse
A database that collects business information from many sources in the enterprise, covering all aspects of the company's processes, products, and customers.

A **data warehouse** is a database that holds business information from many sources in the enterprise, covering all aspects of the company's processes, products, and customers. The data warehouse provides business users with a multidimensional view of the data they need to analyze business conditions. Data warehouses allow managers to *drill down* to get more detail or *roll up* to take detailed data and generate aggregate or summary reports. A data warehouse is designed specifically to support management decision making, not to meet the needs of transaction processing systems. A data warehouse stores historical data that has been extracted from operational systems and external data sources (see Figure 5.17). This operational and external data is "cleaned up" to remove inconsistencies and integrated to create a new information database that is more suitable for business analysis.

Figure 5.17
Elements of a Data Warehouse



Data warehouses typically start out as very large databases, containing millions and even hundreds of millions of data records. As this data is collected from the various production systems, a historical database is built that business analysts can use. To keep it fresh and accurate, the data warehouse receives regular updates. Old data that is no longer needed is purged from the data warehouse. Updating the data warehouse must be fast, efficient, and automated, or the ultimate value of the data warehouse is sacrificed. It is common for a data warehouse to contain from three to ten years of current and historical data. Data-cleaning tools can merge data from many sources into one database, automate data collection and verification, delete unwanted data, and maintain data in a database management system. Data warehouses can also get data from unique sources. Oracle's Warehouse Management software, for example, can accept information from Radio Frequency Identification (RFID) technology, which is being used to tag products as they are shipped or moved from one location to another. Instead of recalling hundreds of thousands of cars because of a possible defective part, automotive companies could determine exactly which cars had the defective parts and only recall the 10,000 cars with the bad parts using RFID. The savings would be huge.



1-800-flowers.com uses a data warehouse to reference customer historical data. DBMS software accessed over the corporate intranet gives marketing professionals the information they need to determine customer interests based on past interactions.[23]

The primary advantage of data warehousing is the ability to relate data in innovative ways. However, a data warehouse can be extremely difficult to establish, with the typical cost exceeding $2 million. Table 5.7 compares online transaction processing (OLTP) and data warehousing.

Data Marts

A **data mart** is a subset of a data warehouse. Data marts bring the data warehouse concept—online analysis of sales, inventory, and other vital business data that has been gathered from transaction processing systems—to small and medium-sized businesses and to departments within larger companies. Rather than store all enterprise data in one monolithic database, data marts contain a subset of the data for a single aspect of a company's business—for example, finance, inventory, or personnel. In fact, a specific area in the data mart might contain more detailed data than the data warehouse would provide.

data mart
A subset of a data warehouse.

Data marts are most useful for smaller groups who want to access detailed data. A warehouse contains summary data that can be used by an entire company. Because data marts typically contain tens of gigabytes of data, as opposed to the hundreds of gigabytes in data warehouses, they can be deployed on less powerful hardware with smaller secondary storage devices, delivering significant savings to an organization. Although any database software can be used to set up a data mart, some vendors deliver specialized software designed and priced specifically for data marts. Already, companies such as Sybase, Software AG, Microsoft, and others have announced products and services that make it easier and cheaper to deploy these scaled-down data warehouses. The selling point: Data marts put targeted business information into the hands of more decision makers. For example, the Defense Acquisition University (DAU), which is responsible for continuing education and career management for employees of the U.S. Department of Defense, uses data marts to provide administrators, instructors, and staff with domain-specific information.[24] A data warehouse is used to combine information from more than 50 disconnected sources, and the DBMS then organizes the information into area-specific data marts, which produce reports accessible through an online dashboard application. The system is estimated to save DAU personnel three to five years of labor.

Characteristic	OLTP Database	Data Warehousing
Purpose	Support transaction processing	Support decision making
Source of data	Business transactions	Multiple files, databases—data internal and external to the firm
Data access allowed users	Read and write	Read only
Primary data access mode	Simple database update and query	Simple and complex database queries with increasing use of data mining to recognize patterns in the data
Primary database model employed	Relational	Relational
Level of detail	Detailed transactions	Often summarized data
Availability of historical data	Very limited—typically a few weeks or months	Multiple years
Update process	Online, ongoing process as transactions are captured	Periodic process, once per week or once per month
Ease of process	Routine and easy	Complex, must combine data from many sources; data must go through a data cleanup process
Data integrity issues	Each transaction must be closely edited	Major effort to "clean" and integrate data from multiple sources

Table 5.7

Comparison of OLTP and Data Warehousing

Data Mining

Data mining is an information-analysis tool that involves the automated discovery of patterns and relationships in a data warehouse. Like gold mining, data mining sifts through mountains of data to find a few nuggets of valuable information. The University of Maryland has developed a data-mining technique to "forecast terrorist behavior based on past actions."[25] The system uses a real-time data extraction tool called T-REX to scour an average of 128,000 articles a day and forecast future activities of over 110 terrorist groups.

data mining
An information-analysis tool that involves the automated discovery of patterns and relationships in a data warehouse.

Data mining's objective is to extract patterns, trends, and rules from data warehouses to evaluate (i.e., predict or score) proposed business strategies, which will improve competitiveness, increase profits, and transform business processes. It is used extensively in marketing to improve customer retention; cross-selling opportunities; campaign management; market, channel, and pricing analysis; and customer segmentation analysis (especially one-to-one marketing). In short, data-mining tools help users find answers to questions they haven't thought to ask.

E-commerce presents another major opportunity for effective use of data mining. Attracting customers to Web sites is tough; keeping them can be next to impossible. For example, when retail Web sites launch deep-discount sales, they cannot easily determine how many first-time customers are likely to come back and buy again. Nor do they have a way of understanding which customers acquired during the sale are price sensitive and more likely to jump on future sales. As a result, companies are gathering data on user traffic through their Web sites and storing the data in databases. This data is then analyzed using data-mining techniques to personalize the Web site and develop sales promotions targeted at specific customers.

predictive analysis
A form of data mining that combines historical data with assumptions about future conditions to predict outcomes of events, such as future product sales or the probability that a customer will default on a loan.

Predictive analysis is a form of data mining that combines historical data with assumptions about future conditions to predict outcomes of events, such as future product sales or the probability that a customer will default on a loan. Retailers use predictive analysis to upgrade occasional customers into frequent purchasers by predicting what products they will buy if offered an appropriate incentive. Genalytics, Magnify, NCR Teradata, SAS Institute,

MySpace.com mines the data of all of its members to determine which ads should be displayed for each member to attract the maximum attention and hits.[26]

Sightward, SPSS, and Quadstone have developed predictive analysis tools. Predictive analysis software can be used to analyze a company's customer list and a year's worth of sales data to find new market segments that could be profitable.

The City of Richmond Police Department uses predictive analysis to predict "when and where crimes were likely to occur, so officers can be on hand to prevent their occurrence."[27]

(Source: Courtesy of Mitch Kezar.)

Traditional DBMS vendors are well aware of the great potential of data mining. Thus, companies such as Oracle, Sybase, Tandem, and Red Brick Systems are all incorporating data-mining functionality into their products. Table 5.8 summarizes a few of the most frequent applications for data mining.

Application	Description
Branding and positioning of products and services	Enable the strategist to visualize the different positions of competitors in a given market using performance (or other) data on dozens of key features of the product and then to condense all that data into a perceptual map of only two or three dimensions.
Customer churn	Predict current customers who are likely to switch to a competitor.
Direct marketing	Identify prospects most likely to respond to a direct marketing campaign (such as a direct mailing).
Fraud detection	Highlight transactions most likely to be deceptive or illegal.
Market basket analysis	Identify products and services that are most commonly purchased at the same time (e.g., nail polish and lipstick).
Market segmentation	Group customers based on who they are or on what they prefer.
Trend analysis	Analyze how key variables (e.g., sales, spending, promotions) vary over time.

Table 5.8

Common Data-Mining Applications

Business Intelligence

The use of databases for business-intelligence purposes is closely linked to the concept of data mining. **Business intelligence** (**BI**) involves gathering enough of the right information in a timely manner and usable form and analyzing it so that it can have a positive effect on business strategy, tactics, or operations. IMS Health, for example, provides a BI system designed to assist businesses in the pharmaceutical industry with custom marketing to physicians, pharmacists, nurses, consumers, government agencies, and nonprofit healthcare organizations.[28] Business intelligence turns data into useful information that is then distributed throughout an enterprise. It provides insight into the causes of problems, and when implemented can improve business operations and sometimes even save lives. For example, BI software at the Sahlgrenska University Hospital in Gothenburg, Sweden, has helped neurosurgeons save lives by identifying complications in patient conditions after cranial surgery.[29] The Information Systems at Work box shows how business intelligence is used in the utilities industry.

business intelligence
The process of gathering enough of the right information in a timely manner and usable form and analyzing it to have a positive impact on business strategy, tactics, or operations.

Competitive intelligence is one aspect of business intelligence and is limited to information about competitors and the ways that knowledge affects strategy, tactics, and operations. Competitive intelligence is a critical part of a company's ability to see and respond quickly and appropriately to the changing marketplace. Competitive intelligence is not espionage—the use of illegal means to gather information. In fact, almost all the information a competitive-intelligence professional needs can be collected by examining published information sources, conducting interviews, and using other legal, ethical methods. Using a variety of analytical tools, a skilled competitive-intelligence professional can by deduction fill the gaps in information already gathered.

competitive intelligence
One aspect of business intelligence limited to information about competitors and the ways that knowledge affects strategy, tactics, and operations.

The term **counterintelligence** describes the steps an organization takes to protect information sought by "hostile" intelligence gatherers. One of the most effective counterintelligence measures is to define "trade secret" information relevant to the company and control its dissemination.

counterintelligence
The steps an organization takes to protect information sought by "hostile" intelligence gatherers.

INFORMATION SYSTEMS @ WORK

Yangtze Power Harnesses the Power

Perhaps you've heard of the Yangtze River in China, and the enormous Three Gorges Dam being erected to harness the river's force for hydroelectric power. Due to be completed in 2011, the Three Gorges Dam will generate 22,500 megawatts of electricity, more than any other hydroelectric facility in the world. The company that will operate the dam is Yangtze Power, China's largest publicly listed utility company.

For years, Yangtze Power has managed the Gezhouba Power Station and six commissioned generating units. It has maintained business data in five databases, supporting its five divisions: Power Generation Management, Finance, Human Resources, Contract Management, and Safety and Control Management. Keeping data in siloed systems—separate, unconnected systems—limited information transfer through the enterprise. If a manager from Human Resources wanted to evaluate data from Contract Management, he would have to e-mail someone in that department to have a report generated and transferred. As Yangtze Power looked ahead to growth and the addition of the world's largest hydroelectric power generator, the company knew that its information would need to flow more freely through the enterprise in order for it to make the best business decisions in a timely fashion.

After evaluating products from Business Objects, Cognos, Informatica, MicroStrategy, and Oracle, Yangtze Power decided to go with Oracle to design one centralized database for all of its information because it was the only company that could provide one integrated system.

In March 2007, Yangtze Power's technology team worked with Oracle to develop a needs analysis and begin data preparation. Requirements were defined to cover six major areas of the business, including 65 performance indices and 370 reports. Through extensive preparation and testing, the system was up and running by November 2007.

Oracle's business intelligence tools allow senior managers to analyze performance on a daily basis, highlight areas for improvement, and monitor the results of business strategies. Each morning, reports on the previous day's critical activities are waiting on managers' desks. The new database stores three years of data, so that managers can draw on historical data when analyzing business performance. Communication between departments has improved, since everyone accesses the same data from a central system, and reports can easily be generated tailored to meet any business need.

Oracle's BI tools are used to create customized reports and charts including pie charts, broken curve diagrams, histograms, and radar maps. Being able to visualize data and trends in data enables a deeper analysis of the organization's business performance.

Yangtze Power has gained control over its flow of information through the enterprise. Now it is working to gain similar control over the raging waters of the Yangtze River.

Discussion Questions

1. What was wrong with Yangtze Power's previous database system, and how was it affecting the business?
2. What solution did Yangtze opt for, and how did it improve business?

Critical Thinking Questions

1. How does a centralized database improve communications within an organization?
2. In what situations might one centralized database not be practical for an enterprise?

Sources: Oracle Success Stories, "Yangtze Power Improves Business Intelligence with Integrated Database and Analysis Tools," 2008, *www.oracle.com/customers/snapshots/yangtze-power-case-study.pdf*, Yangtze River Web site, *www.yangtzeriver.org*, accessed April 2, 2008, Oracle Database and BI Tools, *www.oracle.com/database*, accessed April 2, 2008.

Distributed Databases

distributed database
A database in which the data can be spread across several smaller databases connected via telecommunications devices.

Distributed processing involves placing processing units at different locations and linking them via telecommunications equipment. A **distributed database**—a database in which the data can be spread across several smaller databases connected through telecommunications devices—works on much the same principle. A user in the Milwaukee branch of a clothing manufacturer, for example, might make a request for data that is physically located at corporate headquarters in Milan, Italy. The user does not have to know where the data is physically stored (see Figure 5.18).

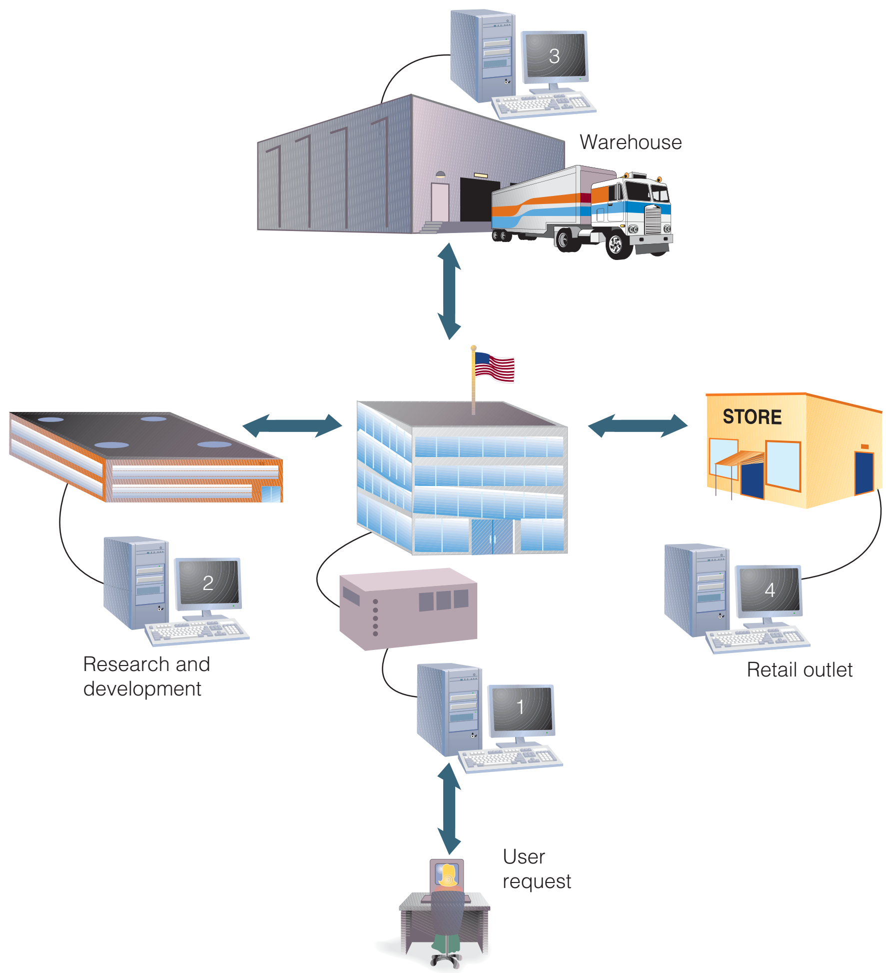


Figure 5.18

The Use of a Distributed Database

For a clothing manufacturer, computers might be located at corporate headquarters, in the research and development center, in the warehouse, and in a company-owned retail store. Telecommunications systems link the computers so that users at all locations can access the same distributed database no matter where the data is actually stored.

Distributed databases give corporations and other organizations more flexibility in how databases are organized and used. Local offices can create, manage, and use their own databases, and people at other offices can access and share the data in the local databases. Giving local sites more direct access to frequently used data can improve organizational effectiveness and efficiency significantly. The New York City Police Department, for example, has thousands of officers searching for information located on servers in offices around the city.

Despite its advantages, distributed processing creates additional challenges in integrating different databases (information integration), maintaining data security, accuracy, timeliness, and conformance to standards. Distributed databases allow more users direct access at different sites; thus, controlling who accesses and changes data is sometimes difficult. Also, because distributed databases rely on telecommunications lines to transport data, access to data can be slower.

To reduce telecommunications costs, some organizations build a replicated database. A **replicated database** holds a duplicate set of frequently used data. The company sends a copy of important data to each distributed processing location when needed or at predetermined times. Each site sends the changed data back to update the main database on an update cycle that meets the needs of the organization. This process, often called *data synchronization*, is used to make sure that replicated databases are accurate, up to date, and consistent with each other. A railroad, for example, can use a replicated database to increase punctuality, safety, and reliability. The primary database can hold data on fares, routings, and other essential information. The data can be continually replicated and downloaded on a read-only basis from the master database to hundreds of remote servers across the country. The remote locations can send back the latest figures on ticket sales and reservations to the main database.

replicated database
A database that holds a duplicate set of frequently used data.

Online Analytical Processing (OLAP)

For nearly two decades, multidimensional databases and their analytical information display systems have provided flashy sales presentations and trade show demonstrations. All you have to do is ask where a certain product is selling well, for example, and a colorful table showing sales performance by region, product type, and time frame appears on the screen. Called **online analytical processing** (**OLAP**), these programs are now being used to store and deliver data warehouse information efficiently. The leading OLAP software vendors include Microsoft, Cognos, SAP, Business Objects, MicroStrategy, Applix, Infor, and Oracle. Lufthansa Cargo depends on OLAP to deliver up-to-the-minute company statistics that help the company compete in the growing global air-freight market.[30] The market is growing by six percent annually, and competitors are emerging all around the world to get a piece of the action. Lufthansa Cargo uses OLAP to analyze its data to provide the fastest service to its customers and the lowest rates.

online analytical processing (OLAP)
Software that allows users to explore data from a number of perspectives.

The value of data ultimately lies in the decisions it enables. Powerful information-analysis tools in areas such as OLAP and data mining, when incorporated into a data warehousing architecture, bring market conditions into sharper focus and help organizations deliver greater competitive value. OLAP provides top-down, query-driven data analysis; data mining provides bottom-up, discovery-driven analysis. OLAP requires repetitive testing of user-originated theories; data mining requires no assumptions and instead identifies facts and conclusions based on patterns discovered. OLAP, or multidimensional analysis, requires a great deal of human ingenuity and interaction with the database to find information in the database. A user of a data-mining tool does not need to figure out what questions to ask; instead, the approach is, “Here’s the data, tell me what interesting patterns emerge.” For example, a data-mining tool in a credit card company’s customer database can construct a profile of fraudulent activity from historical information. Then, this profile can be applied to all incoming transaction data to identify and stop fraudulent behavior, which might otherwise go undetected. Table 5.9 compares the OLAP and data-mining approaches to data analysis.

Table 5.9

Comparison of OLAP and Data Mining

Characteristic	OLAP	Data Mining
Purpose	Supports data analysis and decision making	Supports data analysis and decision making
Type of analysis supported	Top-down, query-driven data analysis	Bottom-up, discovery-driven data analysis
Skills required of user	Must be very knowledgeable of the data and its business context	Must trust in data-mining tools to uncover valid and worthwhile hypotheses

Object-Relational Database Management Systems

object-oriented database
A database that stores both data and its processing instructions.

An **object-oriented database** uses the same overall approach of objected-oriented programming that was discussed in Chapter 4. With this approach, both the data and the processing instructions are stored in the database. For example, an object-oriented database could store monthly expenses and the instructions needed to compute a monthly budget from those expenses. A traditional DBMS might only store the monthly expenses. The King County Metro Transit system in the state of Washington uses an object-oriented database in a system supplied by German vendor Init to manage the routing and accounting of its bus line.[31] Object-oriented databases are useful when a database contains complex data that needs to be processed quickly and efficiently.

In an object-oriented database, a *method* is a procedure or action. A sales tax method, for example, could be the procedure to compute the appropriate sales tax for an order or sale—for example, multiplying the total amount of an order by five percent, if that is the local sales tax. A *message* is a request to execute or run a method. For example, a sales clerk could issue a message to the object-oriented database to compute sales tax for a new order. Many object-oriented databases have their own query language, called *object query language (OQL)*, which is similar to SQL, discussed previously.

object-oriented database management system (OODBMS)
A group of programs that manipulate an object-oriented database and provide a user interface and connections to other application programs.

An object-oriented database uses an **object-oriented database management system (OODBMS)** to provide a user interface and connections to other programs. Computer vendors who sell or lease OODBMSs include Versant and Objectivity. Many organizations are selecting object-oriented databases for their processing power. Versant's OODBMS, for example, is being used by companies in the telecommunications, defense, online gaming, and healthcare industries, and by government agencies. The *Object Data Standard* is a design standard created by the *Object Database Management Group (www.odmg.org)* for developing object-oriented database systems.

object-relational database management system (ORDBMS)
A DBMS capable of manipulating audio, video, and graphical data.

An **object-relational database management system (ORDBMS)** provides a complete set of relational database capabilities plus the ability for third parties to add new data types and operations to the database. These new data types can be audio, images, unstructured text, spatial, or time series data that require new indexing, optimization, and retrieval features. Each of the vendors offering ORDBMS facilities provides a set of application programming interfaces to allow users to attach external data definitions and methods associated with those definitions to the database system. They are essentially offering a standard socket into which users can plug special instructions. DataBlades, Cartridges, and Extenders are the names applied by Oracle and IBM to describe the plug-ins to their respective products. Other plug-ins serve as interfaces to Web servers.

Visual, Audio, and Other Database Systems

In addition to raw data, organizations are increasingly finding a need to store large amounts of visual and audio signals in an organized fashion. Credit card companies, for example, enter pictures of charge slips into an image database using a scanner. The images can be stored in the database and later sorted by customer name, printed, and sent to customers along with their monthly statements. Image databases are also used by physicians to store x-rays and transmit them to clinics away from the main hospital. Financial services, insurance companies, and government branches are using image databases to store vital records and replace paper documents. Drug companies often need to analyze many visual images from laboratories. Chesapeake Energy maintains a database filled with scanned images of terrain and drilling locations.[32] Visual databases can be stored in some object-relational databases or special-purpose database systems. Many relational databases can also store graphic content.

Combining and analyzing data from different databases is an increasingly important challenge. Global businesses, for example, sometimes need to analyze sales and accounting data stored around the world in different database systems. Companies such as IBM are developing *virtual database systems* to allow different databases to work together as a unified database system. Banc of America Securities Prime Brokerage, for example, turned to database virtualization to address management and performance problems. Since its implementation, the virtual database system has reduced storage administration by 95 percent and decreased the need for more storage capacity by 50 percent.[33]

In addition to visual, audio, and virtual databases, other special-purpose database systems meet particular business needs. *Spatial data technology* involves using a database to store and access data according to the locations it describes and to permit spatial queries and analysis. MapInfo software from Pitney Bowes allows businesses such as Home Depot, Sonic Restaurants, CVS Corporation, and Chico's to choose the optimal location for new stores and restaurants based on geospatial demographics.[34] The software provides information about local competition, populations, and traffic patterns to predict how a business will fare in a particular location. Builders and insurance companies use spatial data to make decisions related to natural hazards. Spatial data can even be used to improve financial risk management with information stored by investment type, currency type, interest rates, and time.

Spatial data technology is used by NASA to store data from satellites and Earth stations. Location-specific information can be accessed and compared.

(Source: Courtesy of NASA.)

SUMMARY

Principle

Data management and modeling are key aspects of organizing data and information.

Data is one of the most valuable resources that a firm possesses. It is organized into a hierarchy that builds from the smallest element to the largest. The smallest element is the bit, a binary digit. A byte (a character such as a letter or numeric digit) is made up of eight bits. A group of characters, such as a name or number, is called a field (an object). A collection of related fields is a record; a collection of related records is called a file. The database, at the top of the hierarchy, is an integrated collection of records and files.

An entity is a generalized class of objects for which data is collected, stored, and maintained. An attribute is a characteristic of an entity. Specific values of attributes—called data items—can be found in the fields of the record describing an entity. A data key is a field within a record that is used to identify the record. A primary key uniquely identifies a record, while a secondary key is a field in a record that does not uniquely identify the record.

Traditional file-oriented applications are often characterized by program-data dependence, meaning that they have data organized in a manner that cannot be read by other programs. To address problems of traditional file-based data management, the database approach was developed. Benefits of this approach include reduced data redundancy, improved data consistency and integrity, easier modification and updating, data and program independence, standardization of data access, and more-efficient program development.

One of the tools that database designers use to show the relationships among data is a data model. A data model is a map or diagram of entities and their relationships. Enterprise data modeling involves analyzing the data and information needs of an entire organization. Entity-relationship (ER) diagrams can be employed to show the relationships between entities in the organization.

The relational model places data in two-dimensional tables. Tables can be linked by common data elements, which are used to access data when the database is queried. Each row represents a record. Columns of the tables are called attributes, and allowable values for these attributes are called the domain. Basic data manipulations include selecting, projecting, and joining. The relational model is easier to control, more flexible, and more intuitive than the other models because it organizes data in tables.

Principle

A well-designed and well-managed database is an extremely valuable tool in supporting decision making.

A DBMS is a group of programs used as an interface between a database and its users and other application programs. When an application program requests data from the database, it follows a logical access path. The actual retrieval of the data follows a physical access path. Records can be considered in the same way: A logical record is what the record contains; a physical record is where the record is stored on storage devices. Schemas are used to describe the entire database, its record types, and their relationships to the DBMS.

A DBMS provides four basic functions: providing user views, creating and modifying the database, storing and retrieving data, and manipulating data and generating reports. Schemas are entered into the computer via a data definition language, which describes the data and relationships in a specific database. Another tool used in database management is the data dictionary, which contains detailed descriptions of all data in the database.

After a DBMS has been installed, the database can be accessed, modified, and queried via a data manipulation language. A more specialized data manipulation language is the query language, the most common being Structured Query Language (SQL). SQL is used in several popular database packages today and can be installed on PCs and mainframes.

Popular single-user DBMSs include Corel Paradox and Microsoft Access. IBM, Oracle, and Microsoft are the leading DBMS vendors. Database as a Service (DaaS), or Database 2.0, is a new form of database service in which clients lease use of a database on a service provider's site.

Selecting a DBMS begins by analyzing the information needs of the organization. Important characteristics of databases include the size of the database, the number of concurrent users, its performance, the ability of the DBMS to be integrated with other systems, the features of the DBMS, the vendor considerations, and the cost of the database management system.

Principle

The number and types of database applications will continue to evolve and yield real business benefits.

Traditional online transaction processing (OLTP) systems put data into databases very quickly, reliably, and efficiently, but they do not support the types of data analysis that today's businesses and organizations require. To address this need, organizations are building data warehouses, which are relational database management systems specifically designed to support management decision making. Data marts are subdivisions of data warehouses, which are commonly devoted to specific purposes or functional business areas.

Data mining, which is the automated discovery of patterns and relationships in a data warehouse, is emerging as a practical approach to generating hypotheses about the patterns and anomalies in the data that can be used to predict future behavior.

Predictive analysis is a form of data mining that combines historical data with assumptions about future conditions to forecast outcomes of events such as future product sales or the probability that a customer will default on a loan.

Business intelligence is the process of getting enough of the right information in a timely manner and usable form and analyzing it so that it can have a positive effect on business strategy, tactics, or operations. Competitive intelligence is one aspect of business intelligence limited to information about competitors and the ways that information affects strategy, tactics, and operations. Competitive intelligence is not espionage—the use of illegal means to gather information. Counterintelligence describes the steps an organization takes to protect information sought by "hostile" intelligence gatherers.

With the increased use of telecommunications and networks, distributed databases, which allow multiple users and different sites access to data that may be stored in different physical locations, are gaining in popularity. To reduce telecommunications costs, some organizations build replicated databases, which hold a duplicate set of frequently used data.

Multidimensional databases and online analytical processing (OLAP) programs are being used to store data and allow users to explore the data from a number of different perspectives.

An object-oriented database uses the same overall approach of objected-oriented programming, first discussed in Chapter 4. With this approach, both the data and the processing instructions are stored in the database. An object-relational database management system (ORDBMS) provides a complete set of relational database capabilities, plus the ability for third parties to add new data types and operations to the database. These new data types can be audio, video, and graphical data that require new indexing, optimization, and retrieval features.

In addition to raw data, organizations are increasingly finding a need to store large amounts of visual and audio signals in an organized fashion. A number of special-purpose database systems are also being used.

CHAPTER 5: SELF-ASSESSMENT TEST

Data management and modeling are key aspects of organizing data and information.

1. A group of programs that manipulate the database and provide an interface between the database and the user of the database and other application programs is called a(n) ________________.
   a. GUI
   b. operating system
   c. DBMS
   d. productivity software
2. A(n) ________________ is a skilled and trained IS professional who directs all activities related to an organization's database.
3. Data redundancy is a desirable quality in a database. True or False?
4. A(n) ________________ is a field or set of fields that uniquely identifies a database record.
   a. attribute
   b. data item
   c. key
   d. primary key
5. A(n) ________________ uses basic graphical symbols to show the organization of and relationships between data.
6. What database model places data in two-dimensional tables?
   a. relational
   b. network
   c. normalized
   d. hierarchical

A well-designed and well-managed database is an extremely valuable tool in supporting decision making.

7. ________________ involves combining two or more database tables.
8. After data has been placed into a relational database, users can make inquiries and analyze data. Basic data manipulations include selecting, projecting, and optimizing. True or False?
9. Because the DBMS is responsible for providing access to a database, one of the first steps in installing and using a database involves telling the DBMS the logical and physical structure of the data and relationships among the data in the database. This description of an entire database is called a(n) ________________.

10. The commands used to access and report information from the database are part of the ________________.
    a. data definition language
    b. data manipulation language
    c. data normalization language
    d. schema
11. Access is a popular DBMS for ________________.
    a. personal computers
    b. graphics workstations
    c. mainframe computers
    d. supercomputers
12. A new trend in database management, known as Database as a Service, places the responsibility of storing and managing a database on a service provider. True or False?

The number and types of database applications will continue to evolve and yield real business benefits.

13. A(n) ________________ holds business information from many sources in the enterprise, covering all aspects of the company's processes, products, and customers.
14. An information-analysis tool that involves the automated discovery of patterns and relationships in a data warehouse is called ________________.
    a. a data mart
    b. data mining
    c. predictive analysis
    d. business intelligence
15. ________________ allows users to predict the future based on database information from the past and present.

CHAPTER 5: SELF-ASSESSMENT TEST ANSWERS

(1) c (2) database administrator (3) False (4) d (5) entity-relationship diagram (6) a (7) Joining (8) False (9) schema (10) b (11) a (12) True (13) data warehouse (14) b (15) Predictive analysis

REVIEW QUESTIONS

1. What is an attribute? How is it related to an entity?
2. Define the term *database*. How is it different from a database management system?
3. What is the hierarchy of data in a database?
4. What is a flat file?
5. What is the purpose of a primary key? How can it be useful in controlling data redundancy?
6. What is the purpose of data cleanup?
7. What are the advantages of the database approach?
8. What is data modeling? What is its purpose? Briefly describe three commonly used data models.
9. What is a database schema, and what is its purpose?
10. How can a data dictionary be useful to database administrators and DBMS software engineers?
11. Identify important characteristics in selecting a database management system.
12. What is the difference between a data definition language (DDL) and a data manipulation language (DML)?
13. What is the difference between projecting and joining?
14. What is a distributed database system?
15. What is a data warehouse, and how is it different from a traditional database used to support OLTP?
16. What is meant by the "front end" and the "back end" of a DBMS?
17. What is data mining? What is OLAP? How are they different?
18. What is an ORDBMS? What kind of data can it handle?
19. What is business intelligence? How is it used?
20. In what circumstances might a database administrator consider using an object-oriented database?

DISCUSSION QUESTIONS

1. You have been selected to represent the student body on a project to develop a new student database for your school. What actions might you take to fulfill this responsibility to ensure that the project meets the needs of students and is successful?
2. Your company wants to increase revenues from its existing customers. How can data mining be used to accomplish this objective?
3. You are going to design a database for your cooking club to track its recipes. Identify the database characteristics

most important to you in choosing a DBMS. Which of the database management systems described in this chapter would you choose? Why? Is it important for you to know what sort of computer the database will run on? Why or why not?

4. Make a list of the databases in which data about you exists. How is the data in each database captured? Who updates each database and how often? Is it possible for you to request a printout of the contents of your data record from each database? What data privacy concerns do you have?
5. If you were the database administrator for the iTunes store, how might you use predictive analysis to determine which artists and movies will sell most next year?
6. You are the vice president of information technology for a large, multinational consumer packaged goods company (such as Procter & Gamble or Unilever). You must make a presentation to persuade the board of directors to invest $5 million to establish a competitive-intelligence organization—including people, data-gathering services, and software tools. What key points do you need to make in favor of this investment? What arguments can you anticipate that others might make?
7. Briefly describe how visual and audio databases can be used by companies today.
8. Identity theft, where people steal your personal information, continues to be a threat. Assume that you are the database administrator for a corporation with a large database. What steps would you implement to help prevent people from stealing personal information from the corporate database?
9. What roles do databases play in your favorite online activities and Web sites?

PROBLEM-SOLVING EXERCISES

1. Develop a simple data model for the music you have on your MP3 player or in your CD collection, where each row is a song. For each row, what attributes should you capture? What will be the unique key for the records in your database? Describe how you might use the database.
2. A video movie rental store is using a relational database to store information on movie rentals to answer customer questions. Each entry in the database contains the following items: Movie ID No. (primary key), Movie Title, Year Made, Movie Type, MPAA Rating, Number of Copies on Hand, and Quantity Owned. Movie types are comedy, family, drama, horror, science fiction, and western. MPAA ratings are G, PG, PG-13, R, NC-17, and NR (not rated). Use a single-user database management system to build a data-entry screen to enter this data. Build a small database with at least ten entries.
3. To improve service to their customers, the salespeople at the video rental store have proposed a list of changes being considered for the database in the previous exercise. From this list, choose two database modifications and modify the data-entry screen to capture and store this new information.

   Proposed changes:
   a. Add the date that the movie was first available to help locate the newest releases.
   b. Add the director's name.
   c. Add the names of three primary actors in the movie.
   d. Add a rating of one, two, three, or four stars.
   e. Add the number of Academy Award nominations.
4. Your school maintains information about students in several interconnected database files. The student_contact file contains student contact information. The student_grades file contains student grade records, and the student_financial file contains financial records including tuition and student loans. Draw a diagram of the fields these three files might contain, which field is a primary key in each file, and which fields serve to relate one file to another. Use Figure 5.7 as a guide.

TEAM ACTIVITIES

1. In a group of three or four classmates, communicate with the person at your school that supervises information systems. Find out how many databases are used by your school and for what purpose. Also find out what policies and procedures are in place to protect the data stored from identity thieves and other threats.
2. As a team of three or four classmates, interview business managers from three different businesses that use databases to help them in their work. What data entities and data attributes are contained in each database? How do they access the database to perform analysis? Have they received training in any query or reporting tools? What do they like about their database and what could be improved? Do any of them use data-mining or OLAP techniques? Weighing the information obtained, select one of these databases as

being most strategic for the firm and briefly present your selection and the rationale for the selection to the class.

3. Imagine that you and your classmates are a research team developing an improved process for evaluating auto loan applicants. The goal of the research is to predict which applicants will become delinquent or forfeit their loan. Those who score well on the application will be accepted; those who score exceptionally well will be considered for lower-rate loans. Prepare a brief report for your instructor addressing these questions:

a. What data do you need for each loan applicant?
b. What data might you need that is not typically requested on a loan application form?
c. Where might you get this data?
d. Take a first cut at designing a database for this application. Using the chapter material on designing a database, show the logical structure of the relational tables for this proposed database. In your design, include the data attributes you believe are necessary for this database, and show the primary keys in your tables. Keep the size of the fields and tables as small as possible to minimize required disk drive storage space. Fill in the database tables with the sample data for demonstration purposes (ten records). After your design is complete, implement it using a relational DBMS.

WEB EXERCISES

1. Use a Web search engine to find information on specific products for one of the following topics: business intelligence, object-oriented databases, or database as a service. Write a brief report describing what you found, including a description of the database products and the companies that developed them.
2. List your five favorite Web sites. Consider the services that they provide. For each site, suggest how one or more databases might be used on the back end to supply information to visitors.

CAREER EXERCISES

1. What type of data is stored by businesses in a professional field that interests you? How many databases might be used to store that data? How would the data be organized within each database?
2. How could you use business intelligence (BI) to do a better job at work? Give some specific examples of how BI can give you a competitive advantage.

CASE STUDIES

Case One

The Getty Vocabularies

J. Paul Getty was an American industrialist who made his fortune in the oil business. He made his first million at age 25 in 1916, and later became the world's first billionaire. Getty viewed art as a 'civilizing influence in society, and strongly believed in making art available to the public for its education and enjoyment.' To that end, he created an art museum in Los Angeles, California, and established the J. Paul Getty Trust, commonly referred to as the Getty.

The Getty includes four branches: the Getty Museum, a research institute, a conservation institute, and a foundation. In the 1980s, the Getty discovered a need within the art research community. Researchers lacked a common vocabulary with which to discuss art and artists' work. Establishing a scientific vocabulary with which to describe artwork, style, and technique would allow the study and appreciation of artwork to flourish. To meet this need, the Getty created and published the *Art and Architecture Thesaurus* (AAT) in 1990. The three-volume tome, which includes a thesaurus of geographic names and the Union List of Artist Names, has become a priceless resource for art historical research. It provides tools, standards, and best practices for documenting works of art, just as the Library of Congress provides a standard cataloging tool for libraries.

However, the massive AAT is difficult to search and is expensive to edit and update. Recognizing that a digital version of the resource would provide many benefits, the Getty recently began porting the AAT and associated volumes into a database that can be electronically searched and edited over the Web. To do so, the Getty had to first select a database technology in which to house the information, and a DBMS for use in searching and editing the contents.

One challenge of building an online AAT was that the various components of the resource were stored using different proprietary technologies. The first task was to collect them into one common technology, which required a custom-designed system. Technicians within the Getty opted to use Oracle databases and a product called PowerBuilder from Sybase, Inc., for the user interface. Custom coding was done in Perl and SQR programming languages to merge the components into a cohesive system. The result is a system called the Vocabulary Coordination System (VCS). The VCS is used to collect, analyze, edit, merge, and distribute the terminology managed by the Getty vocabularies. A special Web-based interface was developed that made searching the volumes easy enough for anyone to manage. You can try it yourself at *www.getty.edu/research/conducting_research/vocabularies*.

The resulting system was so impressive that it won the Getty the Computerworld Honors Award in Media, Arts & Entertainment for innovative use of technology. The system makes it easy for scholars to update information in the vocabularies, and for everyone from school children to professional art historians to research and learn about art and art history. The Getty online vocabularies are an ideal realization of J. Paul Getty's original philosophy of promoting human civility through cultural awareness, creativity, and aesthetic enjoyment.

Discussion Questions

1. What purpose do the Getty vocabularies serve, and how are they supported through database technology?
2. How does using the Web as a front end to this database further support J. Paul Getty's vision?

Critical Thinking Questions

1. What concerns do you think the designers of the database had when making this valuable resource available online to the general public?
2. Why did the database designers need to use custom-designed code to collect the original data?

Sources: Pratt, Mary K., "The Getty makes art accessible with online database." *Computerworld*, March 10, 2008, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=Databases&articleId=310236&taxonomyId=173&intsrc=kc_li_story*; Staff, "The Computerworld Honors Program: Web-Based Global Art Resources: The Getty Vocabularies," *Computerworld*, 2007, *www.cwhonors.org/viewCaseStudy.asp?NominationID=112*; The Getty Web site, *www.getty.edu*, accessed April 1, 2008.

Case Two

ETAI Manages Auto Parts Overload with Open-Source Database

If you need a hard-to-find automobile part for a European import, you could probably find it in a catalog published by the ETAI Group in France. The ETAI catalog includes over 30 million parts for over 50,000 European car models manufactured during the past 15 years. The catalog is updated 100 times each year to stay current with the latest models.

While maintaining an average auto parts catalog might not seem a daunting task, this one is an exception. ETAI collects auto parts information from nine databases provided by parts manufacturers. Each database uses a unique design with different formats for parts numbers and varying amounts and types of fields for each part record. Over many years, ETAI had developed a system for collating the data using a variety of programming languages and platforms. The entire process required 15 steps and two to three weeks. It was so complicated that if ETAI's database administrator were to leave, his replacement would have a difficult time learning how the complicated system worked.

Philippe Bobo, the director of software and information systems at ETAI, knew it was time to improve the system. He and his team tested products from a variety of vendors over a five-week period, and eventually decided to work with Talend Open Data Solutions, based in Los Altos, California. Talend specializes in open-source database management systems that integrate data from various types of systems into a single target system—exactly what ETAI needed.

Talend designed a system for ETAI using a single standard programming language that queries the nine auto parts databases and streams the results into one data warehouse. It then cleans the data and standardizes it for output to a catalog format. The 15-step, three-week process is now reduced to one step and two days.

Philippe likes the open-source nature of Talend's solution because it makes it possible for his own software engineers to work with and adjust the software over time to accommodate new needs in the system. Updating the DBMS has reduced labor costs and production time, and made it possible for ETAI to expand into other types of catalogs and service manuals.

Discussion Questions

1. What challenges did ETAI face that made creating their catalog a three-week-long ordeal?
2. How did the solution provided by Talend reduce the job time by 90 percent?

Critical Thinking Questions

1. What benefits were provided by the open-source solution?
2. Why couldn't ETAI standardize the data formats in the nine databases?

Sources: Weiss, Todd R., "ETAI avoids data traffic jam with open source," *Computerworld*, December 17, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=90531 61&intsrc=news_list*; Weiss, Todd R., "ETAI Rides Open Source to Ease Data Traffic Jam," *Computerworld*, December 31, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=30982 1*; Talend Open Data Solutions Web site, *www.talend.com*, accessed March 31, 2008; ETAI Web site, *www.etai.fr/g_instit/atout.htm*, accessed March 31, 2008.

Questions for Web Case

See the Web site for this book to read about the Whitmann Price Consulting case for this chapter. Following are questions concerning this Web case.

Whitmann Price Consulting: Database Systems and Business Intelligence

Discussion Questions

1. How will Whitmann Price consultants and the company itself benefit from their ability to call up corporate information in an instant anywhere and at any time?
2. Why will the database itself not require a change to support the new advanced mobile communications and information system?

Critical Thinking Questions

1. The Web has acted as a convenient standard for accessing all types of information from various types of computing platforms. How will this benefit the systems developers of Whitmann Price in developing forms and reports for the new mobile system?
2. What are the suggested limitations of using a BlackBerry device for accessing and interacting with corporate data?

NOTES

Sources for opening vignette: Havenstein, Heather, "Wal-Mart CTO details HP data warehouse move," *ITWorld Canada*, August 3, 2007, *www.itworldcanada.com/a/Enterprise-Business-Applications/efb96e0a-18de-47e6-ac61-ddab5cc55b5b.html*; Wal-Mart Corporate Fact Sheet, accessed March 30, 2008; Hayes Weier, Mary, "Wal-Mart Speaks Out On HP Neoview Decision," *Information Week*, August 3, 2007, *www.informationweek.com/management/showArticle.jhtml?articleID=201203010, www.walmartstores.com/media/factsheets/fs_2230.pdf*; HP Neoview Enterprise Data Warehouse Web site, *http://h20331.www2.hp.com/enterprise/cache/414444-0-0-225-121.html*, accessed March 30, 3008.

1 Wailgum, Thomas, "Hollywood agency updates systems to woo talent," *Computerworld*, January 19, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=Business_Intelligence&a rticleId=9008545&taxonomyId=9&intsrc=kc_li_story*.

2 Havenstein, Heather, "City of Albuquerque puts BI capabilities into residents' hands," *Computerworld*, September 17, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=Data_Mining&articleId = 301748&taxonomyId=54&intsrc=kc_li_story*.

3 Vijayan, Jaikumar, "Harvard grad students hit in computer intrusion," *Computerworld*, March 13, 2008, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9068221&source=rss_news10*.

4 Staff, "Bacs database fault leaves 400,000 without pay," *Computerworld UK*, March 30, 2007, *www.itworldcanada.com/a/Information-Architecture/94cb5e9e-79c5-4f29-a909-c7025be0d0b4.html*.

5 Mearian, Lucas, "Study: Digital universe and its impact bigger than we thought," *Computerworld*, March 11, 2008, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9067639&source=rss_news10*.

6 Koman, Richard, "Exploding Digital Data Growth Is a Challenge for IT," *Top Tech News*, March 11, 2008, *www.toptechnews.com/story.xhtml?story_id=58752*.

7 Nakashima, Ellen, "FBI Prepares Vast Database Of Biometrics," *Washington Post*, December 22, 2007, *www.washingtonpost.com/wp-dyn/content/article/2007/12/21/AR2007122102544_pf.html*.

8 Kolbasuk McGee, Marianne, "Wal-Mart Requires In-Store Clinics To Use E-Health Records System," *Information Week*, February 16, 2008, *www.informationweek.com/story/showArticle.jhtml?articleID=206504257&cid=RSSfeed_IWK_All*.

9 Pratt, Mary, "Steven Barlow: Master of Data Warehousing," *Computerworld*, July 9, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=data_warehousing&arti cleId=297032&taxonomyId=55&intsrc=kc_feat*.

10 Mullins, Craig, "The Database Report - July 2007," *TDNA*, July 10, 2007, *www.tdan.com/view-featured-columns/5603*.

11 Fonseca, Brian, "Baseball Hall of Fame on deck for archive format change?" *Computerworld*, July 24, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName=servers_and_data_cente r&articleId=9027849&taxonomyId=154&intsrc=kc_top*.

12 Microsoft Staff, "Microsoft Transforms Management Training into an Interactive, On-the-Job Experience," Microsoft Case Studies, August 30, 2007, *www.microsoft.com/casestudies/casestudy.aspx?casestudyid=4000000613*.

13 Oracle Staff, "Lighting Manufacturer Surya Roshni Streamlines Supply Chain with Bright Results," Oracle Customer Snapshot, 2008, *www.oracle.com/customers/snapshots/surya-roshni-snapshot.pdf*.

14 Staff, "INTELLIFIT Moves From Virtual Fitting (match-to-order) to True Mass Customization: Custom-made jeans with a high-tech twist," *Mass Customization & Open Innovation News*, February 15, 2008, *http://mass-customization.blogs.com/mass_customization_open_i/2008/02/intellifit-move.html*.

15 IT Redux Web site, accessed March 23, 2008, *http://itredux.com/office-20/database/?family=Database*.

16 Lai, Eric, "Cloud database vendors: What, us worry about Microsoft?" *Computerworld*, March 12, 2008, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9067979&pageNumber=1*.

17 Intuit Staff, "Jet Blue – The Challenge," Intuit Case Studies, Accessed April 10, 2008, http://quickbase.intuit.com/customers/jetblue_v4.pdf

18 Arellano, Nestor, "Information created in 2007 will exceed storage capacity says IDC," *ITWorld Canada*, March 23, 2008, *www.itworldcanada.com/a/Information-Architecture/844a38f1-ccaf-445f-867f-498f041587c1.html*.

19 Gittlen, Sandra, "Data center land grab: How to get ready for the rush," *Computerworld*, March 12, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9012963*.

20 AP Staff, "Google ventures into health records biz," CNN.com, February 21, 2008, *www.cnn.com/2008/TECH/02/21/google.records.ap*.

21 Dixon, Guy, "Yahoo embraces semantic Web standards," vnunet.com, March 18, 2008, *www.vnunet.com/vnunet/news/2212249/yahoo-embraces-semantic-web*.

22 McMillan, Robert, "Researcher: Half a million database servers have no firewall," *Computerworld*, November 14, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9046821&source=rss_news10*.

23 SAS Staff, "1-800-FLOWERS.COM Gathers a Bouquet of CRM Capabilities," SAS Success Stories, *www.sas.com/success/1800flowers_IT.html*, accessed March 23, 2008.

24 Computerworld Staff, "The Computerworld Honors Program: Defense Acquisition University," *Computerworld*, 2007, *www.cwhonors.org/viewCaseStudy.asp?NominationID=313*.

25 Vijayan, Jaikumar, "Univ. of Md. launches data mining portal for counter-terrorism research," *Computerworld*, February 26, 2008, *www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9064938*.

26 Stone, Brad, "MySpace to Discuss Effort to Customize Ads," *New York Times*, September 18, 2007, *www.nytimes.com/2007/09/18/technology/18myspace.html?_r=1&oref=slogin*.

27 Computerworld Staff, "The Computerworld Honors Program: City of Richmond Police Department," *Computerworld*, 2007, *www.cwhonors.org/viewCaseStudy.asp?NominationID=303*.

28 Nickum, Chris, "The BI Prescription," *OptimizeMag.com*, April 2007, Page 45

29 Computerworld Staff, "The Computerworld Honors Program: QlikTech International," *Computerworld*, 2007, *www.cwhonors.org/viewCaseStudy.asp?NominationID=155*.

30 Cognos Staff, "Lufthansa Cargo," Lufthansa Cargo Case Study, Accessed April 10, 2008, http://www.cognos.com/pdfs/success_stories/ss_lufthansa-cargo.pdf

31 Staff, "init wins contract worth 25 mill. US Dollar from Seattle," Init News Release, March 13, 2007, *www.init-ka.de/en_news/PR_AH_2007/AH_070313_Seattle_en.php*.

32 Babcock, Charles, "Oracle Touts Virtualization, Apps – and 11g Of Course," *Information Week*, November 19, 2007, page 30.

33 Fonseca, Brian, "Virtualization Cutting Storage Costs for Some Large Firms," *Computerworld*, April 23, 2007, *www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyId=154&articleId=290230&int src=hm_topic*.

34 Staff, "MapInfo Customer Testimonials," MapInfo Web site, *www.mapinfo.com/location/integration?txtTopNav=836a2545d8a37f00dev-vcm100001a031dc7____&txtLeftNav=6e5819cd57f47f00dev-vcm100001a031dc7____&txtExtNav=391495ac2a771110Vgn-VCM10000021021dc7____*, accessed March 23, 2008.