STUDENT LEARNING OBJECTIVES

After completing this chapter, you will be able to answer the following questions:

1. How does a relational database organize data and how does it differ from an object-oriented database?
2. What are the principles of a database management system?
3. What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?
4. What is the role of information policy and data administration in the management of organizational data resources?
5. Why is data quality assurance so important for a business?
The National Association for Stock Car Auto Racing, better known as NASCAR, is on its way to becoming America’s most popular spectator sport. In 2005, 75 million people attended NASCAR races, the highest number of attendees of any sport in the United States. NASCAR racing is second only to the National Football League (NFL) in television ratings. From 1995 to 2004, NASCAR’s fan base ballooned from 63 million to 75 million, and retail sales of NASCAR-licensed merchandise jumped more than 250 percent, from approximately $600 million to $2.1 billion.

What’s wrong with this picture? Not much, except NASCAR management thinks the company can still do better. It thinks it can grow NASCAR’s fan base even more, especially in areas outside the South where NASCAR has traditionally been most popular. But NASCAR needs the right data to identify who its fans are, and it is having trouble putting all those data together.

Until about five years ago, NASCAR considered information systems a back-office function. Its entire race process—obtaining credentials for sponsors, running races, managing relationships with sponsors, and paying drivers and their
teams—was heavily manual and paper-oriented. Race tracks did not have technology to transmit race data effectively.

The picture changed after NASCAR started recruiting top-notch executives from Fortune 500, sports, and media companies who appreciated what IT could do for a business. Roger Lovell, NASCAR’s Managing Director of IT, was empowered to launch an IT-enabled business transformation.

Lovell and his team of information systems specialists developed a scalable, stable IT infrastructure, standardized NASCAR’s desktop computing technology, and implemented a plan to beef up systems security. They began gathering more data electronically about driver performance during a race and analyzing the data so fans could follow drivers’ performance similar to the way baseball uses batting averages. In 2006, NASCAR implemented a state-of-the-art Mobile Technology Center to collect and process timing and scoring data as races take place.

Now Lovell and Roger VanDerSnick, NASCAR’s vice president of marketing, are working on building a single comprehensive database of racing car fans that could be shared with business partners and used for marketing to those fans. Such a database provides a better understanding of NASCAR fans’ demographic profiles, behaviors, and preferences for targeting offers such as NASCAR-branded merchandise, opportunities to meet drivers at races, or products and services from sponsors. In addition to generating revenue, these initiatives would increase fan loyalty, help NASCAR attract new sponsors, and convince existing sponsors to increase their spending.

Creating a consolidated fan database is challenging because the data are housed in many disparate databases. NASCAR.com has its own database as does the NASCAR members club and each driver’s fan club. It is likely that these databases do not store fan data in a consistent manner, so the data will need to be “cleansed” for discrepancies, inconsistencies, and errors, and restructured to fit a single standard format. NASCAR and its partners will have to establish rules for using a consolidated fan database so that NASCAR fans are not bombarded with excessive marketing appeals and the privacy agreements that racing teams, drivers, and sponsors have established with their own fans are respected. Stoneacre Partners, who built a relational database for the Official NASCAR Membership Club, is in charge of the project.


NASCAR’s experience illustrates the importance of data management and database systems for business. NASCAR has experienced phenomenal growth. But its future growth and business performance depend on what it can or can’t do with its customer data. How businesses store, organize, and manage their data has a tremendous impact on organizational effectiveness.

The chapter-opening diagram calls attention to important points raised by this case and this chapter. Management decided that NASCAR’s business strategy needed to focus on creating customer intimacy. Data about NASCAR fans and potential customers had been stored in a number of different databases where they could not be easily retrieved and analyzed. NASCAR did not have all the information it needed for identifying all of its existing and potential fans to market its products and services.

In the past, NASCAR had used heavily manual paper processes to manage its information. This solution was no longer viable as the organization grew larger. A more appropriate solution was to integrate NASCAR customer data from all of its disparate sources into a single comprehensive fan database. In addition to using appropriate technology, NASCAR had to correct and reorganize the data into a standard format and establish rules with its business partners for accessing information in the new database.

A comprehensive fan database helps NASCAR and its partners boost profitability by making it easier to target potential customers for their products and services. The database
This chapter focuses on databases and how businesses use databases to achieve their objectives. Small and large companies alike use databases to record business transactions, control inventories, manage employees, and achieve customer and supplier intimacy. Once these data are properly organized in database management systems, the data can be analyzed, and the resulting information can be used to develop new businesses, achieve operational excellence, inform management decision making, and help the firm fulfill its reporting requirements to higher authorities. Entire businesses, such as UPS, credit card companies, and Google, are based on databases. It would not be an overstatement to say that databases are the foundation of business today and that most businesses would fail should their databases cease to exist.

**HEADS UP**

This chapter focuses on databases and how businesses use databases to achieve their objectives. Small and large companies alike use databases to record business transactions, control inventories, manage employees, and achieve customer and supplier intimacy. Once these data are properly organized in database management systems, the data can be analyzed, and the resulting information can be used to develop new businesses, achieve operational excellence, inform management decision making, and help the firm fulfill its reporting requirements to higher authorities. Entire businesses, such as UPS, credit card companies, and Google, are based on databases. It would not be an overstatement to say that databases are the foundation of business today and that most businesses would fail should their databases cease to exist.

### 5.1 The Database Approach to Data Management

A database is a collection of related files containing records on people, places, or things. One of the most successful databases in modern history is the telephone book. The telephone book is a collection of records on people and businesses who use telephones. The telephone book lists four pieces of information for each phone user: last name, first name, address, and phone number. It also contains information on businesses and business categories, such as auto dealers or plumbing suppliers. The telephone book draws its information from a database with files for customers, business classifications, and area codes and geographic regions.
Prior to the development of digital databases, a business would use large filing cabinets filled with paper files to store information on transactions, customers, suppliers, inventory, and employees. They would also use lists, laboriously collated and typed by hand, to quickly summarize the information in paper files. You can still find paper-based manual databases in most doctors’ offices where patient records are stored in thousands of paper files.

Needless to say, paper-based databases are extremely inefficient and costly to maintain, often contain inaccurate data, are slow, and make it difficult to access the data in a timely fashion. Paper-based databases are also extremely inflexible. For instance, it would be nearly impossible for a paper-based doctor’s office to combine its files on prescriptions with its files on patients in order to produce a list of all people for whom they had prescribed a specific drug. For a modern computer database, this would be very easy. In fact, a powerful feature of computer databases is the ability to quickly relate one set of files to another.

**ENTITIES AND ATTRIBUTES**

How do you start thinking about the data for your business and how to manage them? If you are starting up or running a business, you will have to identify the data you will need to run your business. Typically, you will be using data on categories of information, such as customers, suppliers, employees, orders, products, shippers, and perhaps parts. Each of these generalized categories representing a person, place, or thing on which we store and maintain information is called an entity. Each entity has specific characteristics, called attributes. For example, the entity SUPPLIER has specific attributes, such as the supplier’s name and address, which would most likely include street, city, state, and ZIP code. The entity PART typically has attributes such as part description, price of each part (unit price), and supplier who produced the part.

**ORGANIZING DATA IN A RELATIONAL DATABASE**

If you stored this information in paper files, you would probably have a file on each entity and its attributes. In an information system, a database organizes the data much the same way, grouping related pieces of data together. The relational database is the most common type of database today. Relational databases organize data into two-dimensional tables (called relations) with columns and rows. Each table contains data on an entity and its attributes. For the most part, there is one table for each business entity. So, at the most basic level, you will have one table for customers, and a table each for suppliers, parts in inventory, employees, and sales transactions.

Let’s look at how a relational database would organize data about suppliers and parts. Take the SUPPLIER table, which is illustrated in Figure 5-1. It consists of a grid of columns and rows of data. Each individual element of data about a supplier, such as the supplier name,
street, city, state, and ZIP code, is stored as a separate field within the SUPPLIER table. Each field represents an attribute for the entity SUPPLIER. Fields in a relational database are also called columns.

The actual information about a single supplier that resides in a table is called a row. Rows are commonly referred to as records, or, in very technical terms, as tuples.

Note that there is a field for Supplier_Number in this table. This field uniquely identifies each record so that the record can be retrieved, updated, or sorted, and it is called a key field. Each table in a relational database has one field that is designated as its primary key. This key field is the unique identifier for all the information in any row of the table, and this primary key cannot be duplicated.

We could use the supplier’s name as a key field. However, if two different suppliers had the same name (which does happen from time to time), supplier name would not uniquely identify each, so it is necessary to assign a special identifier field for this purpose. For example, if you had two suppliers, both named “CBM,” but one was based in Dayton and another in St. Louis, it would be easy to confuse them. However, if each has a unique Supplier_Number, such confusion is prevented.

We also see that the address information has been separated into four separate fields: Supplier_Street, Supplier_City, Supplier_State, and Supplier_Zip. Data are separated into the smallest elements that one would want to access separately to make it easy to select only the rows in the table that match the contents of one field, such as all the suppliers in Ohio (OH). The rows of data can also be sorted by the contents of the Supplier_State field to get a list of suppliers by state regardless of their cities.

So far, the SUPPLIER table does not have any information about the parts that a particular supplier provides for your company. PART is a separate entity from SUPPLIER, and fields with information about parts should be stored in a separate PART table (see Figure 5-2).

Why not keep information on parts in the same table as suppliers? If we did that, each row of the table would contain the attributes of both PART and SUPPLIER. Because one supplier could supply more than one part, the table would need many extra rows for a single supplier to show all the parts that supplier provided. We would be maintaining a great deal of redundant data about suppliers, and it would be difficult to search for the information on any individual part because you would not know whether this part is the first or fiftieth part in this supplier’s record. A separate table, PART, should be created to store these three fields and solve this problem.

The PART table would also have to contain another field, Supplier_Number, so that you would know the supplier for each part. It would not be necessary to keep repeating all the

---

**Figure 5-2**

The PART Table

Data for the entity PART have their own separate table. Part_Number is the primary key and Supplier_Number is the foreign key, enabling users to find related information from the SUPPLIER table about the supplier for each part.
information about a supplier in each PART record because having a Supplier_Number field in the PART table allows you to “look up” the data in the fields of the SUPPLIER table.

Notice that Supplier_Number appears in both the SUPPLIER and PART tables. In the SUPPLIER table, Supplier_Number is the primary key. When the field Supplier_Number appears in the PART table it is called a foreign key and is essentially a look-up field to look up data about the supplier of a specific part. Note that the PART table would itself have its own primary key field, Part_Number, to uniquely identify each part. This key is not used to link PART with SUPPLIER but might be used to link PART with a different entity.

As we organize data into tables, it is important to make sure that all the attributes for a particular entity apply only to that entity. If you were to keep the supplier’s address with the PART record, that information would not really relate only to PART; it would relate to both PART and SUPPLIER. If the supplier’s address were to change, it would be necessary to alter the data in every PART record rather than only once in the SUPPLIER record.

ESTABLISHING RELATIONSHIPS

Now that we’ve broken down our data into a SUPPLIER table and a PART table, we must make sure we understand the relationship between them. A schematic called an entity-relationship diagram is used to clarify table relationships in a relational database. The most important piece of information provided by an entity-relationship diagram is the manner in which two tables are related to each other. Tables in a relational database may have one-to-one, one-to-many, and many-to-many relationships.

An example of a one-to-one relationship might be a situation where a human resources system must store confidential data about employees. It might store data, such as the employee name, date of birth, address, and job position in one table, and confidential data about that employee, such as salary or pension benefits, in another table. These two tables pertaining to a single employee would have a one-to-one relationship because each record in the EMPLOYEE table with basic employee data has only one related record in the table storing confidential data.

The relationship between the SUPPLIER and PART entities in our database is a one-to-many relationship: Each supplier can supply more than one part, but each part has only one supplier. For every record in the SUPPLIER table, there may be many related records in the PART table.

Figure 5-3 illustrates how an entity-relationship diagram would depict this one-to-many relationship. The boxes represent entities. The lines connecting the boxes represent relationships. A line connecting two entities that ends in two short marks designates a one-to-one relationship. A line connecting two entities that ends with a crow’s foot topped by a short mark indicates a one-to-many relationship. Figure 5-3 shows that each PART has only one SUPPLIER, but many PARTs can be provided by the same SUPPLIER.

We would also see a one-to-many relationship if we wanted to add a table about orders to our database because one supplier services many orders. The ORDER table would only contain the Order_Number and Order_Date. Figure 5-4 illustrates a report showing an order of parts from a supplier. If you look at the report, you can see that the information on the top-right portion of the report comes from the ORDER table. The actual line items ordered are listed in the lower portion of the report.

Figure 5-3
A Simple Entity-Relationship Diagram
This diagram shows the relationship between the entities SUPPLIER and PART.
Because one order can be for many parts from a supplier and a single part can be ordered many times on different orders, this creates a many-to-many relationship between the PART and ORDER tables. Whenever a many-to-many relationship exists between two tables, it is necessary to link these two tables in a table that joins this information. Creating a separate table for a line item in the order would serve this purpose. This table is often called a join table or an intersection relation. This join table contains only three fields: Order_Number and Part_Number, which are used only to link the ORDER and PART tables, and Part_Quantity. If you look at the bottom-left part of the report, this is the information coming from the LINE_ITEM table.

We would thus wind up with a total of four tables in our database. Figure 5-5 illustrates the final set of tables, and Figure 5-6 shows what the entity-relationship diagram for this set of tables would look like. Note that the ORDER table does not contain data on the extended price because that value could be calculated by multiplying Unit_Price by Part_Quantity. This data element can be derived when needed using information that already exists in the PART and LINE_ITEM tables. Order Total is another derived field calculated by totaling the extended prices for items ordered.

The process of streamlining complex groups of data to minimize redundant data elements and awkward many-to-many relationships, and increase stability and flexibility is called normalization. A database that has been properly designed and normalized will be easy to maintain, and will minimize duplicate data. The Learning Tracks at the end of this chapter direct you to more detailed discussions of normalization, entity-relationship diagramming, and database design on the Laudon Web site.

Relational database systems try to enforce referential integrity rules to ensure that relationships between coupled tables remain consistent. When one table has a foreign key that points to another table, you may not add a record to the table with the foreign key unless there is a corresponding record in the linked table. In the database we have just created, the foreign key Supplier_Number links the PART table to the SUPPLIER table. We may not add a new record to the PART table for a part with supplier number 8266 unless there is a corresponding record in the SUPPLIER table for supplier number 8266. We must also delete the corresponding record in the PART table if we delete the record in the SUPPLIER table for supplier number 8266. In other words, we shouldn’t have parts from nonexistent suppliers!

The example provided here for parts, orders, and suppliers is a simple one. Even in a very small business, you will have tables for other important entities, such as customers, shippers, and employees. A very large corporation might have databases with thousands of entities (tables) to maintain. What is important for any business, large or small, is to have a good data model that includes all of its entities and the relationships among them, one that is organized to minimize redundancy, maximize accuracy, and make data easily accessible for reporting and analysis.
The final design of the database for suppliers, parts, and orders has four tables. The LINE_ITEM table is a join table that eliminates the many-to-many relationship between ORDER and PART.

**FIGURE 5-5**
The Final Database Design with Sample Records

The Part Number table includes the following parts:
- **137**: Door latch
- **145**: Side mirror
- **150**: Door molding
- **152**: Door lock
- **155**: Compressor
- **178**: Door handle

The Supplier table includes the following suppliers:
- **8259**: CBM Inc.
- **8261**: B. R. Molds
- **8263**: Jackson Components
- **8444**: Bryant Corporation

The Supplier Street table includes the following addresses:
- **74 5th Avenue**: Dayton, OH 45220
- **1277 Gandolly Street**: Cleveland, OH 49345
- **8233 Micklin Street**: Lexington, KY 56723
- **4315 Mill Drive**: Rochester, NY 11344

The Order table includes the following orders:
- **3502**: 1/15/2008
- **3503**: 1/16/2008
- **3504**: 1/17/2008

The LINE_ITEM table includes the following sample records:
- **3502**: Part Number 137, Part Quantity 10
- **3502**: Part Number 152, Part Quantity 20
- **3502**: Part Number 178, Part Quantity 5
It cannot be emphasized enough: If the business does not get its data model right, the system will not be able to serve the business right. The company’s systems will not be as effective as they could be because they will have to work with data that may be inaccurate, incomplete, or difficult to retrieve. Understanding the organization’s data and how they should be represented in a database is perhaps the most important lesson you can learn from this course.

For example, Famous Footwear, a shoe store chain with more than 800 locations in 49 states, could not achieve its goal of having “the right style of shoe in the right store for sale at the right price” because its database was not properly designed for rapidly adjusting store inventory. The company had an Oracle relational database running on an IBM AS/400 midrange computer, but the database was designed primarily for producing standard reports for management rather than for reacting to marketplace changes. Management could not obtain precise data on specific items in inventory in each of its stores. The company had to work around this problem by building a new database where the sales and inventory data could be better organized for analysis and inventory management.

5.2 Database Management Systems

Now that you have started creating the files and identifying the data required by your business, you will need a database management system to help you manage and use the data. A database management system (DBMS) is a specific type of software for creating, storing, organizing, and accessing data from a database. Microsoft Access is a DBMS for desktop systems, whereas DB2, Oracle Database, and Microsoft SQL Server are DBMS for large mainframes and midrange computers. MySQL is a popular open-source DBMS, and Oracle Database Lite is a DBMS for small handheld computing devices. All of these products are relational DBMS that support a relational database.

The DBMS relieves the end user or programmer from the task of understanding where and how the data are actually stored by separating the logical and physical views of the data. The logical view presents data as end users or business specialists would perceive them, whereas the physical view shows how data are actually organized and structured on physical storage media, such as a hard disk.

The database management software makes the physical database available for different logical views required by users. For example, for the human resources database illustrated in Figure 5-7, a benefits specialist might require a view consisting of the employee’s name, social security number, and health insurance coverage. A payroll department member might need data such as the employee’s name, social security number, gross pay, and net pay. The data for all of these views is stored in a single database, where it can be more easily managed by the organization.

Figure 5-6
Entity-Relationship Diagram for the Database with Four Tables
This diagram shows the relationship between the entities SUPPLIER, PART, LINE_ITEM, and ORDER.
OPERATIONS OF A RELATIONAL DBMS

In a relational database, tables can be easily combined to deliver data required by users, provided that any two tables share a common data element. Let’s return to the database we set up earlier with PART and SUPPLIER tables illustrated in Figures 5-1 and 5-2.

Suppose we wanted to find in this database the names of suppliers who could provide us with part number 137 or part number 150. We would need information from two tables: the SUPPLIER table and the PART table. Note that these two tables have a shared data element: Supplier_Number.

In a relational database, three basic operations, as shown in Figure 5-8, are used to develop useful sets of data: select, project, and join. The select operation creates a subset consisting of all records in the file that meet stated criteria. Select creates, in other words, a subset of rows that meet certain criteria. In our example, we want to select records (rows) from the PART table where the Part_Number equals 137 or 150. The join operation combines relational tables to provide the user with more information than is available in individual tables. In our example, we want to join the now-shortened PART table (only parts 137 or 150 will be presented) and the SUPPLIER table into a single new table.

The project operation creates a subset consisting of columns in a table, permitting the user to create new tables that contain only the information required. In our example, we want to extract from the new table only the following columns: Part_Number, Part_Name, Supplier_Number, and Supplier_Name (see Figure 5-8).

CAPABILITIES OF DATABASE MANAGEMENT SYSTEMS

A DBMS includes capabilities and tools for organizing, managing, and accessing the data in the database. The most important are its data definition capability, data dictionary, and data manipulation language.

DBMS have a data definition capability to specify the structure of the content of the database. It would be used to create database tables and to define the characteristics of the fields in each table. This information about the database would be documented in a data dictionary. A data dictionary is an automated or manual file that stores definitions of data elements and their characteristics. Microsoft Access has a rudimentary data dictionary capability that displays information about the name, description, size, type, format, and other properties of each field in a table (see Figure 5-9). Data dictionaries for large corporate databases may capture additional information, such as usage; ownership (who in the
### Figure 5-8
The Three Basic Operations of a Relational DBMS

The select, project, and join operations enable data from two different tables to be combined and only selected attributes to be displayed.

#### PART

<table>
<thead>
<tr>
<th>Part_Number</th>
<th>Part_Name</th>
<th>Unit_Price</th>
<th>Supplier_Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>Door latch</td>
<td>22.00</td>
<td>8259</td>
</tr>
<tr>
<td>145</td>
<td>Side mirror</td>
<td>12.00</td>
<td>8444</td>
</tr>
<tr>
<td>150</td>
<td>Door molding</td>
<td>6.00</td>
<td>8263</td>
</tr>
<tr>
<td>152</td>
<td>Door lock</td>
<td>31.00</td>
<td>8263</td>
</tr>
<tr>
<td>155</td>
<td>Compressor</td>
<td>54.00</td>
<td>8261</td>
</tr>
<tr>
<td>176</td>
<td>Door handle</td>
<td>10.00</td>
<td>8259</td>
</tr>
</tbody>
</table>

#### SUPPLIER

<table>
<thead>
<tr>
<th>Supplier_Number</th>
<th>Supplier_Name</th>
<th>Supplier_Street</th>
<th>Supplier_City</th>
<th>Supplier_State</th>
<th>Supplier_Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>8259</td>
<td>CBM Inc.</td>
<td>74 5th Avenue</td>
<td>Dayton</td>
<td>OH</td>
<td>45220</td>
</tr>
<tr>
<td>8261</td>
<td>B. R. Molds</td>
<td>1277 Gandolly Street</td>
<td>Cleveland</td>
<td>OH</td>
<td>49345</td>
</tr>
<tr>
<td>8263</td>
<td>Jackson Components</td>
<td>8233 Micklin Street</td>
<td>Lexington</td>
<td>KY</td>
<td>56723</td>
</tr>
<tr>
<td>8444</td>
<td>Bryant Corporation</td>
<td>4315 Mill Drive</td>
<td>Rochester</td>
<td>NY</td>
<td>11344</td>
</tr>
</tbody>
</table>

**Select** Part_Number = 137 or 150

**Join by Supplier_Number**

**Project selected columns**

<table>
<thead>
<tr>
<th>Part_Number</th>
<th>Part_Name</th>
<th>Supplier_Number</th>
<th>Supplier_Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>Door latch</td>
<td>8259</td>
<td>CBM Inc.</td>
</tr>
<tr>
<td>150</td>
<td>Door molding</td>
<td>8263</td>
<td>Jackson Components</td>
</tr>
</tbody>
</table>
organization is responsible for maintaining the data); authorization; security; and the individuals, business functions, programs, and reports that use each data element.

**Querying and Reporting**

DBMS include tools for accessing and manipulating information in databases. Most DBMS have a specialized language called a data manipulation language that is used to add, change, delete, and retrieve the data in the database. This language contains commands that permit end users and programming specialists to extract data from the database to satisfy information requests and develop applications. The most prominent data manipulation language today is **Structured Query Language**, or **SQL**. Figure 5-10 illustrates the SQL query that would produce the new resultant table in Figure 5-8. You can find out more about how to perform SQL queries in our Learning Tracks for this chapter, which can be found on the Laudon Web site.

Users of DBMS for large and midrange computers, such as DB2, Oracle, or SQL Server, would employ SQL to retrieve information they needed from the database. Microsoft Access also uses SQL, but it provides its own set of user-friendly tools for querying databases and for organizing data from databases into more polished reports.

Microsoft Access has capabilities to help users create queries by identifying the tables and fields they want and the results, and then selecting the rows from the database that meet

**Figure 5-9**

**Access Data**

**Dictionary Features**

Microsoft Access has a rudimentary data dictionary capability that displays information about the size, format, and other characteristics of each field in a database. Displayed here is the information maintained in the SUPPLIER table. The small key icon to the left of Supplier_Number indicates that it is a key field.

**Figure 5-10**

**Example of an SQL Query**

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a list with the same results as Figure 5-8.

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number
SUPPLIER.Supplier_Name
FROM PART, SUPPLIER
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND
Part_Number = 137 OR Part_Number = 150;
```
particular criteria. These actions in turn are translated into SQL commands. Figure 5-11 illustrates how the same query as the SQL query to select parts and suppliers in Figure 5-10 would be constructed using Microsoft Access.

DBMS typically include capabilities for report generation so that the data of interest can be displayed in a more structured and polished format than would be possible just by querying. Crystal Reports is a popular report generator for large corporate DBMS, although it can also be used with Microsoft Access.

Microsoft Access also has capabilities for developing desktop system applications. These include tools for creating data entry screens, reports, and developing the logic for processing transactions. These capabilities are primarily used by information systems specialists.

OBJECT-ORIENTED DATABASES

Many applications today and in the future require databases that can store and retrieve not only structured numbers and characters but also drawings, images, photographs, voice, and full-motion video. DBMS designed for organizing structured data into rows and columns are not well suited to handling graphics-based or multimedia applications. Object-oriented databases are better suited for this purpose.

An object-oriented DBMS stores the data and procedures that act on those data as objects that can be automatically retrieved and shared. Object-oriented database management systems (OODBMS) are becoming popular because they can be used to manage the various multimedia components or Java applets used in Web applications, which typically integrate pieces of information from a variety of sources.

Although object-oriented databases can store more complex types of information than relational DBMS, they are relatively slow compared with relational DBMS for processing large numbers of transactions. Hybrid object-relational DBMS systems are now available to provide capabilities of both object-oriented and relational DBMS.
5.3 Using Databases to Improve Business Performance and Decision Making

Businesses use their databases to keep track of basic transactions, such as paying suppliers, processing orders, serving customers, and paying employees. But they also need databases to provide information that will help the company run the business more efficiently, and help managers and employees make better decisions. If a company wants to know which product is the most popular or who is its most profitable customer, the answer lies in the data.

For example, by analyzing data from customer credit card purchases, Louise’s Trattoria, a Los Angeles restaurant chain, learned that quality was more important than price for most of its customers, who were college-educated and liked fine wine. Acting on this information, the chain introduced vegetarian dishes, more seafood selections, and more expensive wines, raising sales by more than 10 percent.

In a large company, with large databases or large systems for separate functions, such as manufacturing, sales, and accounting, special capabilities and tools are required for analyzing vast quantities of data and for accessing data from multiple systems. These capabilities include data warehousing, data mining, and tools for accessing internal databases through the Web.

**DATA WAREHOUSES**

What if you wanted concise, reliable information about current operations, trends, and changes across the entire company? If you worked in a large company, this might be difficult because data are often maintained in separate systems, such as sales, manufacturing, or accounting. Some of the data you needed might be found in the sales system, and other pieces in the manufacturing system. Many of these systems are older legacy systems that use outdated data management technologies or file systems where information is difficult for users to access.

You might have to spend an inordinate amount of time locating and gathering the data you needed, or you would be forced to make your decision based on incomplete knowledge. If you wanted information about trends, you might also have trouble finding data about past events because most firms only make their current data immediately available. Data warehousing addresses these problems.

**WHAT IS A DATA WAREHOUSE?**

A data warehouse is a database that stores current and historical data of potential interest to decision makers throughout the company. The data originate in many core operational transaction systems, such as systems for sales, customer accounts, and manufacturing, and may include data from Web site transactions. The data warehouse consolidates and standardizes information from different operational databases so that the information can be used across the enterprise for management analysis and decision making.

Figure 5-12 illustrates how a data warehouse works. The data warehouse makes the data available for anyone to access as needed, but it cannot be altered. A data warehouse system also provides a range of ad hoc and standardized query tools, analytical tools, and graphical reporting facilities. Many firms use intranet portals to make the data warehouse information widely available throughout the firm.

**DATA MARTS**

Companies often build enterprise-wide data warehouses, where a central data warehouse serves the entire organization, or they create smaller, decentralized warehouses called data marts. A data mart is a subset of a data warehouse in which a summarized or highly focused portion of the organization’s data is placed in a separate database for a specific population of users. For example, a company might develop marketing and sales data marts to
deal with customer information. A data mart typically focuses on a single subject area or line of business, so it usually can be constructed more rapidly and at lower cost than an enterprise-wide data warehouse.

**BUSINESS INTELLIGENCE, MULTIDIMENSIONAL DATA ANALYSIS, AND DATA MINING**

Once data have been captured and organized in data warehouses and data marts, they are available for further analysis. A series of tools enables users to analyze these data to see new patterns, relationships, and insights that are useful for guiding decision making. These tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions are often referred to as **business intelligence (BI)**. Principal tools for business intelligence include software for database querying and reporting, tools for multidimensional data analysis (online analytical processing), and data mining tools.

When we think of intelligence as applied to humans, we typically think of people’s ability to combine learned knowledge with new information and change behaviors in such a way that they succeed at their task or adapt to a new situation. Likewise, business intelligence provides firms with the capability to amass information; develop knowledge about customers, competitors, and internal operations; and change decision-making behavior to achieve higher profitability and other business goals.

For instance, Harrah’s Entertainment, the second-largest gambling company in its industry, continually analyzes data about its customers gathered when people play its slot machines or use Harrah’s casinos and hotels. Harrah’s marketing department uses this information to build a detailed gambling profile, based on a particular customer’s ongoing value to the company. For instance, business intelligence lets Harrah’s know the favorite gaming experience of a regular customer at one of its Midwest riverboat casinos, along with that person’s preferences for room accommodations, restaurants, and entertainment. This information guides management decisions about how to cultivate the most profitable customers, encourage those customers to spend more, and attract more customers with high revenue-generating potential. Business intelligence has improved Harrah’s profits so much that it has become the centerpiece of the firm’s business strategy.
Figure 5-13 illustrates how business intelligence works. The firm’s operational databases keep track of the transactions generated by running the business. These databases feed data to the data warehouse. Managers use business intelligence tools to find patterns and meanings in the data. Managers then act on what they have learned from analyzing the data by making more informed and intelligent business decisions.

This section will introduce you to the most important business intelligence technologies and tools. We will provide more detail about business intelligence applications in the Chapter 10 discussion of decision making.

**Online Analytical Processing (OLAP)**
Suppose your company sells four different products—nuts, bolts, washers, and screws—in the East, West, and Central regions. If you wanted to ask a fairly straightforward question, such as how many washers sold during the past quarter, you could easily find the answer by querying your sales database. But what if you wanted to know how many washers sold in each of your sales regions and compare actual results with projected sales?

To obtain the answer, you would need **online analytical processing (OLAP)**. OLAP supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions. Each aspect of information—product, pricing, cost, region, or time period—represents a different dimension. So, a product manager could use a multidimensional data analysis tool to learn how many washers were sold in the East in June, how that compares with the previous month and the previous June, and how it compares with the sales forecast. OLAP enables users to obtain online answers to ad hoc questions such as these in a fairly rapid amount of time, even when the data are stored in very large databases, such as sales figures for multiple years.

Figure 5-14 shows a multidimensional model that could be created to represent products, regions, actual sales, and projected sales. A matrix of actual sales can be stacked on top of a matrix of projected sales to form a cube with six faces. If you rotate the cube 90 degrees one way, the face showing will be product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. If you rotate 180 degrees from the original view, you will see projected sales and product versus region. Cubes can be nested within cubes to build complex views of data. A company would use either a specialized multidimensional database or a tool that creates multidimensional views of data in relational databases.
DATA MINING

Traditional database queries answer such questions as, “How many units of product number 403 were shipped in February 2008?” OLAP, or multidimensional analysis, supports much more complex requests for information, such as, “Compare sales of product 403 relative to plan by quarter and sales region for the past two years.” With OLAP and query-oriented data analysis, users need to have a good idea about the information for which they are looking.

Data mining is more discovery driven. Data mining provides insights into corporate data that cannot be obtained with OLAP by finding hidden patterns and relationships in large databases and inferring rules from them to predict future behavior. The patterns and rules are used to guide decision making and forecast the effect of those decisions. The types of information obtainable from data mining include associations, sequences, classifications, clusters, and forecasts.

• Associations are occurrences linked to a single event. For instance, a study of supermarket purchasing patterns might reveal that, when corn chips are purchased, a cola drink is purchased 65 percent of the time, but when there is a promotion, cola is purchased 85 percent of the time. This information helps managers make better decisions because they have learned the profitability of a promotion.

• In sequences, events are linked over time. We might find, for example, that if a house is purchased, a new refrigerator will be purchased within two weeks 65 percent of the time, and an oven will be bought within one month of the home purchase 45 percent of the time.

• Classification recognizes patterns that describe the group to which an item belongs by examining existing items that have been classified and by inferring a set of rules. For example, businesses such as credit card or telephone companies worry about the loss of steady customers. Classification helps discover the characteristics of customers who are likely to leave and can provide a model to help managers predict who those customers are so that the managers can devise special campaigns to retain such customers.

• Clustering works in a manner similar to classification when no groups have yet been defined. A data mining tool can discover different groupings within data, such as finding affinity groups for bank cards or partitioning a database into groups of customers based on demographics and types of personal investments.

• Although these applications involve predictions, forecasting uses predictions in a different way. It uses a series of existing values to forecast what other values will be.

Figure 5-14
Multidimensional Data Model

The view that is showing is product versus region. If you rotate the cube 90 degrees, the face that will show is product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. Other views are possible.
For example, forecasting might find patterns in data to help managers estimate the future value of continuous variables, such as sales figures.

These systems perform high-level analyses of patterns or trends, but they can also drill down to provide more detail when needed. There are data mining applications for all the functional areas of business, and for government and scientific work. One popular use for data mining is to provide detailed analyses of patterns in customer data for one-to-one marketing campaigns or for identifying profitable customers.

For example, Virgin Mobile Australia uses a data warehouse and data mining to increase customer loyalty and roll out new services. The company created a data warehouse that consolidated data from its enterprise system, customer relationship management system, and customer billing systems in a massive database. Data mining has enabled management to determine the demographic profile of new customers and relate it to the handsets they purchased as well as to the performance of each store and point-of-sale campaign, consumer reactions to new products and services, customer attrition rates, and the revenue generated by each customer.

**Predictive analysis** uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events, such as the probability a customer will respond to an offer or purchase a specific product. For example, the U.S. division of The Body Shop International plc used predictive analysis with its database of catalog, Web, and retail store customers to identify customers who were more likely to make catalog purchases. That information helped the company build a more precise and targeted mailing list for its catalogs, improving the response rate for catalog mailings and catalog revenues.

Data mining is both a powerful and profitable tool, but it poses challenges to the protection of individual privacy. Data mining technology can combine information from many diverse sources to create a detailed “data image” about each of us—our income, our driving habits, our hobbies, our families, and our political interests. The question of whether companies should be allowed to collect such detailed information about individuals is discussed in Chapter 12. The Interactive Session on Management explores the debate about whether large databases housing DNA profiles used in crime-fighting pose a threat to privacy and social well-being.

**DATABASES AND THE WEB**

Many companies are using the Web to make some of the information in their internal databases available to customers and business partners. Prospective customers might use a company’s Web site to view the company’s product catalog or to place an order. The company in turn might use the Web to check inventory availability for that product from its supplier. That supplier in turn may have to check with its own suppliers as well as delivery firms needed to ship the products on time.

These actions involve accessing and (in the case of ordering) updating corporate databases through the Web. Suppose, for example, a customer with a Web browser wants to search an online retailer’s database for pricing information. Figure 5-15 illustrates how that customer might access the retailer’s internal database over the Web. The user would access
INTERACTIVE SESSION: MANAGEMENT DNA Databases: Crime Fighting Weapon or Threat to Privacy?

On December 2, 1996, Cherie Morrisette, age 11, left her apartment on the south side of Jacksonville, Florida after arguing with her sister about washing the dishes. Six days later, her body was found 13 miles away in the St. Johns River. Detectives classified the case as a homicide, but could not locate a suspect for 10 years. Then on March 16, 2006, detectives in Colchester, Connecticut announced that Robert Shelton Mitchell, age 43, of New Britain, Connecticut would be charged with the murder.

A DNA profile of bodily fluid left by the murderer at the crime scene had been developed by the Biology Laboratory of the Florida Department of Law Enforcement and sent to the nationwide DNA bank. There it matched a DNA sample provided by Mr. Mitchell to the Connecticut sex offenders’ registry in 2003 while he was in prison on another charge.

DNA evidence has become a potent crime-fighting tool, allowing a criminal to be identified by his or her own genes. Computer analysis can discover the identity of a criminal by matching DNA from blood, hair, saliva, or other bodily fluids left at a crime scene with a DNA profile in a database. A laboratory creates a profile of specific agreed-upon genetic segments of the DNA molecule for a specific individual and stores that information in a database. To identify a particular individual, the laboratory compares the profile produced from a sample of unknown DNA with the profile produced from a sample belonging to an identified individual to see if there is a match.

Law enforcement agencies around the world are increasingly relying on DNA evidence. U.S. law enforcement agencies use databases of DNA profiles created by the states and linked through the FBI’s Combined DNA Index System (CODIS). The CODIS system, authorized by Congress in 1994, allows law enforcement officials to exchange and compare DNA profiles at the local, state, and national levels. As of April 2007, CODIS had data on more than 4.6 million profiles. The samples on which the DNA profiles are based, primarily blood or saliva, are kept at forensic laboratories around the country. All 50 states, the FBI, and the U.S. Army work with this system.

The CODIS system has helped law enforcement officials identify suspects in more than 11,000 cases. It helped solve two “cold” murders in Kansas, identify two-decade-old remains of a missing California child, and capture a predator terrorizing young boys in Houston.

DNA identification is also helpful in proving innocence. Lawyer Barry Scheck’s Innocence Project at the Cardozo School of Law in Yeshiva University has used DNA identification to free more than 200 people who had been wrongly convicted. Law schools at the University of Wisconsin, the University of Washington, and Santa Clara University have similar innocence projects.

According to Joseph M. Polisar, president of the International Association of Chiefs of Police, DNA testing “is the fingerprint technology of this century... The potential for us in the criminal justice field to solve crimes with this technology is boundless.”

Despite all their benefits, DNA databases remain controversial. Privacy advocates and defense lawyers believe genetic databases pose risks to the innocent if they contain data on people who are not convicted criminals. In some instances, DNA has been collected from witnesses or others to eliminate them from police inquiries. DNA has also been collected from families of suspects to determine whether suspects should continue to be pursued. The Justice Department is completing rules to allow the collection of DNA from anyone arrested or detained by federal authorities, including illegal immigrants detained by federal agents. Some state legislators have also advocated expanding the FBI national DNA database to include juveniles.

Most people aren’t violent criminals, including those who commit misdemeanors, and their inclusion in a national DNA database exposes them to risks they would not otherwise face. People who collect and analyze DNA can make mistakes. (Sloppy DNA collection and laboratory procedures resulted in at least one wrongful conviction in Houston and may have affected the outcome of the O.J. Simpson trial.) There may be valid reasons for an innocent person’s DNA to be at a crime scene that police might choose to disregard. Innocent people may be caught up in a criminal investigation when their DNA from a single hair or spot of saliva on a drinking glass appears in a public or private place where they had every right to be.

Critics have also pointed out that expanding the FBI’s DNA database would create large backlogs in the FBI's laboratory that logs, analyzes, and stores federal DNA samples. The lab might have to process an additional 250,000 to 1 million samples per year. There has already been an enormous increase in DNA samples to be processed but no new resources for the FBI’s laboratory.

the retailer’s Web site over the Internet using Web browser software on his or her client PC. The user’s Web browser software would request data from the organization’s database, using HTML commands to communicate with the Web server.

Because many “back-end” databases cannot interpret commands written in HTML, the Web server would pass these requests for data to software that translates HTML commands into SQL so that they can be processed by the DBMS working with the database. In a client/server environment, the DBMS resides on a dedicated computer called a database server. The DBMS receives the SQL requests and provides the required data. The information is transferred from the organization’s internal database back to the Web server for delivery in the form of a Web page to the user.

Figure 5-15 shows that the software working between the Web server and the DBMS could be an application server running on its own dedicated computer (see Chapter 4). The application server software handles all application operations, including transaction processing and data access, between browser-based computers and a company’s back-end business applications or databases. The application server takes requests from the Web server, runs the business logic to process transactions based on those requests, and provides connectivity to the organization’s back-end systems or databases. Alternatively, the software for handling these operations could be a custom program or a CGI script. A CGI script is a compact program using the Common Gateway Interface (CGI) specification for processing data on a Web server.

There are a number of advantages to using the Web to access an organization’s internal databases. First, everyone knows how to use Web browser software, and employees require much less training than if they used proprietary query tools. Second, the Web interface requires few or no changes to the internal database. Companies leverage their investments in older systems because it costs much less to add a Web interface in front of a legacy system than to redesign and rebuild the system to improve user access. For this reason, most large Fortune 500 firms all have back-end legacy databases running on mainframe computers that are linked to “front-end” software that makes the information available in the form of a Web page to users on request.

Accessing corporate databases through the Web is creating new efficiencies and opportunities, and, in some cases, it is even changing the way business is being done. ThomasNet.com provides an up-to-date directory of information from more than 700,000 suppliers of industrial products, such as chemicals, metals, plastics, rubber, and automotive equipment. Formerly called Thomas Register, the company used to send out huge paper catalogs with this information. Now, it provides this information to users online via its Web site and has become a smaller, leaner company.
INTERACTIVE SESSION: TECHNOLOGY

The Databases Behind MySpace

MySpace.com, the popular social networking site, has experienced one of the greatest growth spurts in the history of the Internet. The site launched in November 2003 and by May 2007, it had 175 million member accounts. The challenge for MySpace has been to avoid technological letdowns that degrade Web site performance and frustrate its rapidly expanding network of users.

The technical requirements of a site like MySpace are vastly different from other heavily trafficked Web sites. Generally, a small number of people change the content on a news site a few times a day. The site may retrieve thousands of read-only requests from its underlying database without having to update the database. On MySpace, tens of millions of users are constantly updating their content, resulting in an elevated percentage of database interactions that require updates to the underlying database. Each time a user views a profile on MySpace, the resulting page is stitched together from database lookups that organize information from multiple tables stored in multiple databases residing on multiple servers.

In its initial phases, MySpace operated with two Web servers communicating with one database server and a Microsoft SQL Server database. Such a setup is ideal for small to medium-size sites because of its simplicity. At MySpace, the setup showed signs of stress as more users came aboard. At first, MySpace reduced the load by adding Web servers to handle the increased user requests. But when the number of accounts stretched to 500,000 in 2004, one database server was not sufficient. Deploying additional database servers is more complicated than adding Web servers because the data must be divided among multiple databases without any loss in accessibility or performance. MySpace deployed three SQL Server databases. One served as a master database, which received all new data and copied them to the other two databases. These databases focused on retrieving data for user page requests.

As MySpace approached 2 million accounts, the database servers approached their input/output capacity, which refers to the speed at which they could read and write data. This caused the site to lag behind in content updates. MySpace switched to a virtualized storage architecture, which ended the practice of attaching disks dedicated to specific applications in favor of a single pool of storage space available to all applications. Under this arrangement, databases could write data to any available disk, thus eliminating the possibility of an application’s dedicated disk becoming overloaded.

In 2005, MySpace also fortified its infrastructure by installing a layer of servers between the database servers and the Web servers to store and serve copies of frequently accessed data objects so that the site’s Web servers wouldn’t have to query the database servers with lookups as frequently.

Despite all these measures, MySpace still overloads more frequently than other major Web sites. Users have expressed frustration at not being able to log in or view certain pages. Log-in errors occur at a rate of 20 to 40 percent some days. Site activity continues to challenge the limitations of the technology. So far, the site’s continued growth suggests that users are willing to put up with periodic “Unexpected Error” screens. MySpace developers continue to redesign the Web site’s database, software, and storage systems to keep pace with its exploding growth, but their job is never done.

Other companies have created entirely new businesses based on access to large databases through the Web. One is the social networking site MySpace, which helps users stay connected with each other or meet new people. MySpace features music, comedy, videos, and “profiles” with information supplied by 175 million users about their age, hometown, dating preferences, marital status, and interests. It maintains a massive database to house and manage all of this content. Because the site grew so rapidly, it had to make a series of changes to its underlying database technology. The Interactive Session on Technology explores this topic. As you read this case, try to identify the problem MySpace is facing; what alternative solutions are available to management; and the people, organization, and technology issues that have to be addressed when developing the solution.

5.4 Managing Data Resources

Setting up a database is only a start. In order to make sure that the data for your business remain accurate, reliable, and readily available to those who need it, your business will need special policies and procedures for data management.

ESTABLISHING AN INFORMATION POLICY

Every business, large and small, needs an information policy. Your firm’s data are an important resource, and you don’t want people doing whatever they want with them. You need to have rules on how the data are to be organized and maintained, and who is allowed to view the data or change them.

An information policy specifies the organization’s rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying information. Information policies lay out specific procedures and accountabilities, identifying which users and organizational units can share information, where information can be distributed, and who is responsible for updating and maintaining the information. For example, a typical information policy would specify that only selected members of the payroll and human resources department would have the right to change and view sensitive employee data, such as an employee’s salary or social security number, and that these departments are responsible for making sure that such employee data are accurate.

If you are in a small business, the information policy would be established and implemented by the owners or managers. In a large organization, managing and planning for
information as a corporate resource often requires a formal data administration function. **Data administration** is responsible for the specific policies and procedures through which data can be managed as an organizational resource. These responsibilities include developing information policy, planning for data, overseeing logical database design and data dictionary development, and monitoring how information systems specialists and end-user groups use data.

A large organization will also have a database design and management group within the corporate information systems division that is responsible for defining and organizing the structure and content of the database, and maintaining the database. In close cooperation with users, the design group establishes the physical database, the logical relations among elements, and the access rules and security procedures. The functions it performs are called **database administration**.

**ENSURING DATA QUALITY**

A well-designed database and information policy will go a long way toward ensuring that the business has the information it needs. However, additional steps must be taken to ensure that the data in organizational databases are accurate and remain reliable.

What would happen if a customer’s telephone number or account balance were incorrect? What would be the impact if the database had the wrong price for the product you sold? Data that are inaccurate, untimely, or inconsistent with other sources of information create serious operational and financial problems for businesses. When faulty data go unnoticed, they often lead to incorrect decisions, product recalls, and even financial losses.

According to Forrester Research, 20 percent of U.S. mail and commercial package deliveries were returned because of incorrect names or addresses. The Gartner Group consultants reported that more than 25 percent of the critical data in large Fortune 1000 companies’ databases is inaccurate or incomplete, including bad product codes and product descriptions, faulty inventory descriptions, erroneous financial data, incorrect supplier information, and incorrect employee data. Gartner believes that customer data degrades at a rate of 2 percent per month, making poor data quality a major obstacle to successful customer relationship management (Gage and McCormick, 2005).

Some of these data quality problems are caused by redundant and inconsistent data produced by multiple systems feeding a data warehouse. For example, the sales ordering system and the inventory management system might both maintain data on the organization’s products. However, the sales ordering system might use the term *Item Number* and the inventory system might call the same attribute *Product Number*. The sales, inventory, or manufacturing systems of a clothing retailer might use different codes to represent values for an attribute. One system might represent clothing size as “extra large,” whereas the other system might use the code “XL” for the same purpose. During the design process for the warehouse database, data describing entities, such as a customer, product, or order, should be named and defined consistently for all business areas using the database.

If a database is properly designed and enterprise-wide data standards established, duplicate or inconsistent data elements should be minimal. Most data quality problems, however, such as misspelled names, transposed numbers, or incorrect or missing codes, stem from errors during data input. The incidence of such errors is rising as companies move their businesses to the Web and allow customers and suppliers to enter data into their Web sites that directly update internal systems.

Think of all the times you have received several pieces of the same direct mail advertising on the same day. This is very likely the result of having your name maintained multiple times in a database. Your name may have been misspelled or you used your middle initial on one occasion and not on another or the information was initially entered onto a paper form and not scanned properly into the system. Because of these inconsistencies, the
database would treat you as different people! We often receive redundant mail addressed to Laudon, Lavdon, Lauden, or Landon.

Before a new database is in place, organizations need to identify and correct their faulty data and establish better routines for editing data once their database is in operation. Analysis of data quality often begins with a data quality audit, which is a structured survey of the accuracy and level of completeness of the data in an information system. Data quality audits can be performed by surveying entire data files, surveying samples from data files, or surveying end users for their perceptions of data quality.

Data cleansing, also known as data scrubbing, consists of activities for detecting and correcting data in a database that are incorrect, incomplete, improperly formatted, or redundant. Data cleansing not only corrects data but also enforces consistency among different sets of data that originated in separate information systems. Specialized data-cleansing software is available to automatically survey data files, correct errors in the data, and integrate the data in a consistent company-wide format.

5.5 Hands-On MIS

The projects in this section give you hands-on experience in redesigning a customer database for targeted marketing, creating a database for inventory management, and using the Web to search online databases for overseas business resources.

IMPROVING DECISION MAKING: REDESIGNING THE CUSTOMER DATABASE

Software skills: Database design, querying, and reporting
Business skills: Customer profiling

Dirt Bikes USA sells primarily through its distributors. It maintains a small customer database with the following data: customer name, address (street, city, state, ZIP code), telephone number, model purchased, date of purchase, and distributor. The database is illustrated below and you can find it on the Laudon Web site for Chapter 5. These data are collected by its distributors when they make a sale and are then forwarded to Dirt Bikes. Dirt Bikes would like to be able to market more aggressively to its customers.

The marketing department would like to be able to send customers e-mail notices of special racing events and of sales on parts. It would also like to learn more about customers’ interests and tastes: their ages, years of schooling, another sport in which they are interested, and whether they attend dirt bike racing events. Additionally, Dirt Bikes would like to know whether customers own more than one motorcycle. (Some Dirt Bikes customers own two or three motorcycles purchased from Dirt Bikes USA or other manufacturers.) If a motorcycle was purchased from Dirt Bikes, the company would like to know the date of purchase, model purchased, and distributor. If the customer owns a non-Dirt Bikes motorcycle, the company would like to know the manufacturer and model of the other motorcycle (or motorcycles) and the distributor from whom the customer purchased that motorcycle.
• Redesign Dirt Bikes’s customer database so that it can store and provide the information needed for marketing. You will need to develop a design for the new customer database and then implement that design using database software. Consider using multiple tables in your new design. Populate each new table with 10 records.

• Develop several reports that would be of great interest to Dirt Bikes’s marketing and sales department (for example, lists of repeat Dirt Bikes customers, Dirt Bikes customers who attend racing events, or the average ages and years of schooling of Dirt Bikes customers) and print them.

ACHIEVING OPERATIONAL EXCELLENCE: BUILDING A RELATIONAL DATABASE FOR INVENTORY MANAGEMENT

Software skills: Database design, querying, and reporting
Business skills: Inventory management

In this exercise, you will use database software to design a database for managing inventory for a small business. Sylvester’s Bike Shop, located in San Francisco, California, sells road, mountain, hybrid, leisure, and children’s bicycles. Currently, Sylvester’s purchases bikes from three suppliers, but plans to add new suppliers in the near future. This rapidly growing business needs a database system to manage this information.

Initially, the database should house information about suppliers and products. The database will contain two tables: a supplier table and a product table. The reorder level refers to the number of items in inventory that triggers a decision to order more items to prevent a stockout. (In other words, if the number of units of a particular item in inventory falls below the reorder level, the item should be reordered.) The user should be able to perform several queries and produce several managerial reports based on the data contained in the two tables.

Using the information found in the tables on the Laudon Web site for Chapter 5, build a simple relational database for Sylvester’s. Once you have built the database, perform the following activities.

• Prepare a report that identifies the five most expensive bicycles. The report should list the bicycles in descending order from most expensive to least expensive, the quantity on hand for each, and the markup percentage for each.

• Prepare a report that lists each supplier, its products, the quantities on hand, and associated reorder levels. The report should be sorted alphabetically by supplier. Within each supplier category, the products should be sorted alphabetically.

• Prepare a report listing only the bicycles that are low in stock and need to be reordered. The report should provide supplier information for the items identified.

• Write a brief description of how the database could be enhanced to further improve management of the business. What tables or fields should be added? What additional reports would be useful?

IMPROVING DECISION MAKING: SEARCHING ONLINE DATABASES FOR OVERSEAS BUSINESS RESOURCES

Software skills: Online databases
Business skills: Researching services for overseas operations

Internet users have access to many thousands of Web-enabled databases with information on services and products in faraway locations. This project develops skills in searching these online databases.

Your company is located in Greensboro, North Carolina, and manufactures office furniture of various types. You have recently acquired several new customers in Australia, and a study you commissioned indicates that, with a presence there, you could greatly increase your sales. Moreover, your study indicates that you could do even better if you
actually manufactured many of your products locally (in Australia). First, you need to set up an office in Melbourne to establish a presence, and then you need to begin importing from the United States. You then can plan to start producing locally.

You will soon be traveling to the area to make plans to actually set up an office, and you want to meet with organizations that can help you with your operation. You will need to engage people or organizations that offer many services necessary for you to open your office, including lawyers, accountants, import-export experts, telecommunications equipment and support, and even trainers who can help you to prepare your future employees to work for you. Start by searching for U.S. Department of Commerce advice on doing business in Australia. Then try the following online databases to locate companies that you would like to meet with during your upcoming trip: Australian Business Register (abr.gov.au), Australia Trade Now (australiatradenow.com), and the Nationwide Business Directory of Australia (www.nationwide.com.au). If necessary, you could also try search engines such as Yahoo! and Google.

- List the companies you would contact to interview on your trip to determine whether they can help you with these and any other functions you think vital to establishing your office.
- Rate the databases you used for accuracy of name, completeness, ease of use, and general helpfulness.
- What does this exercise tell you about the design of databases?

**LEARNING TRACKS**

The following Learning Tracks provide content relevant to topics covered in this chapter:

1. Database Design, Normalization, and Entity-Relationship Diagramming
2. Introduction to SQL

**Review Summary**

1. **How does a relational database organize data and how does it differ from an object-oriented database?** The relational database is the primary method for organizing and maintaining data today in information system. It organizes data in two-dimensional tables with rows and columns called relations. Each table contains data about an entity and its attributes. Each row represents a record and each column represents an attribute or field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation. An entity-relationship diagram graphically depicts the relationship between entities (tables) in a relational database. The process of breaking down complex groupings of data and streamlining them to minimize redundancy and awkward many-to-many relationships is called normalization. An object-oriented DBMS stores data and procedures that act on the data as objects, and it can handle multimedia as well as characters and numbers.

2. **What are the principles of a database management system?** A database management system (DBMS) consists of software that permits centralization of data and data management so that businesses have a single consistent source for all their data needs. A single database services multiple applications. The DBMS separates the logical and physical views of data so that the user does not have to be concerned with its physical location. The principal capabilities of a DBMS includes a data definition capability, a data dictionary capability, and a data manipulation language.
What are the principal tools and technologies for accessing information from databases to improve business performance and decision making? A data warehouse consolidates current and historical data from many different operational systems in a central database designed for reporting and analysis. Data warehouses support multidimensional data analysis, also known as online analytical processing (OLAP). OLAP represents relationships among data as a multidimensional structure, which can be visualized as cubes of data and cubes within cubes of data. Data mining analyzes large pools of data, including the contents of data warehouses, to find patterns and rules that can be used to predict future behavior and guide decision making. Conventional databases can be linked to the Web or a Web interface to facilitate user access to an organization’s internal data.

What is the role of information policy and data administration in the management of organizational data resources? Developing a database environment requires policies and procedures for managing organizational data as well as a good data model and database technology. A formal information policy governs the maintenance, distribution, and use of information in the organization. In large corporations, a formal data administration function is responsible for information policy, as well as for data planning, data dictionary development, and monitoring data usage in the firm.

Why is data quality assurance so important for a business? Data that are inaccurate, incomplete, or inconsistent create serious operational and financial problems for businesses if they lead to inaccurate decisions about the actions that should be taken by the firm. Assuring data quality involves using enterprise-wide data standards, databases designed to minimize inconsistent and redundant data, data quality audits, and data cleansing software.

### Key Terms

- Attributes, 160
- Business intelligence (BI), 171
- Data administration, 179
- Data cleaning, 180
- Data definition language, 166
- Data dictionary, 166
- Data manipulation language, 168
- Data mart, 170
- Data mining, 173
- Data quality audit, 180
- Data warehouse, 170
- Database, 159
- Database administration, 179
- Database management system (DBMS), 165
- Database server, 176
- Entity, 160
- Entity-relationship diagram, 162
- Field, 161
- Foreign key, 162
- Information policy, 178
- Key field, 161
- Normalization, 163
- Object-oriented DBMS, 169
- Object-relational DBMS, 169
- Online analytical processing (OLAP), 172
- Predictive analysis, 174
- Primary key, 161
- Records, 161
- Referential integrity, 163
- Relational database, 160
- Structured Query Language (SQL), 168
- Tuples, 161

### Review Questions

1. How does a relational database organize data and how does it differ from an object-oriented database?
   - Define and explain the significance of entities, attributes, and key fields.
   - Define a relational database and explain how it organizes and stores information.
   - Explain the role of entity-relationship diagrams and normalization in database design.
   - Define an object-oriented database and explain how it differs from a relational database.
2. What are the principles of a database management system?
   • Define a database management system (DBMS) and describe how it works and its benefits to organizations.
   • Define and compare the logical and a physical view of data.
   • Define and describe the three operations of a relational database management system.
   • Name and describe the three major capabilities of a DBMS.

3. What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?
   • Define a data warehouse and describe how it works.
   • Define business intelligence and explain how it is related to database technology.
   • Describe the capabilities of online analytical processing (OLAP).
   • Define data mining, describe what types of information can be obtained from it, and explain how does it differs from OLAP.
   • Explain how users can access information from a company’s internal databases through the Web.

4. What is the role of information policy and data administration in the management of organizational data resources?
   • Define information policy and data administration and explain how they help organizations manage their data.

5. Why is data quality assurance so important for a business?
   • List and describe the most common data quality problems.
   • List and describe the most important tools and techniques for assuring data quality.

Discussion Questions

1. It has been said that you do not need database management software to create a database environment. Discuss.
2. To what extent should end users be involved in the selection of a database management system and database design?

Video Case

You will find a video case illustrating some of the concepts in this chapter on the Laudon Web site along with questions to help you analyze the case.

Teamwork

Identifying Entities and Attributes in an Online Database

With a group of two or three of your fellow students, select an online database to explore, such as AOL Music or the Internet Movie Database. Explore these Web sites to see what information they provide. Then list the entities and attributes that they must keep track of in their databases. If possible, diagram the relationship between the entities you have identified. If possible, use electronic presentation software to present your findings to the class.
BUSINESS PROBLEM-SOLVING CASE

Can HP Mine Success from an Enterprise Data Warehouse?

Hewlett-Packard, the $98.5 billion manufacturer of personal computers, server computers, printers, and provider of consulting services, is in the middle of a business transformation. The company is trying to reduce yearly spending by its information technology department by 30 percent over five years. HP expects information technology expenditures to drop from just over $3 billion in 2003 to $2.1 billion in 2008. HP is reducing its information system applications from 5,000 to 1,500 and consolidating 85 computer centers to six. HP’s current IT infrastructure employs between 19,000 and 22,000 servers. The consolidation will decrease the total by 8,000 to 9,000.

The success of HP’s business transformation may hinge on one particular project. HP is building a 400-terabyte data warehouse to serve the entire enterprise. If successful, the data warehouse will dispose of 17 different database technologies and unite 14,000 databases currently in use. If the initiative fails, HP would join a long list of organizations that have been confounded by the complexity of implementing enterprise-wide databases.

From an internal perspective, HP’s data warehouse aims to give its workforce access to data in real time with no departmental or geographic boundaries. HP’s numerous systems and applications had serious data management problems. CEO Mark Hurd had difficulty collecting and analyzing “consistent, timely data spanning different parts of the business.” Some systems tracked sales and pricing by product, while others tracked sales information geographically. Commonly used financial information, such as gross margins to measure profitability, were calculated differently from business unit to business unit. The company was obtaining information from more than 750 data marts.

Lack of data consistency dragged down sales and profits. Compiling information about the business from various systems could take up to a week, so managers had to make decisions based on relatively stale data. Seemingly simple questions, such as how much the company was spending on marketing across its different businesses, were difficult to answer. Without a consistent view of the enterprise, senior executives struggled with decisions on matters such as the size of sales and service teams assigned to particular systems.

HP CIO Randy Mott began consolidating the data marts in November 2005 into a single data warehouse serving the entire enterprise. He created a team composed of 300 people who were running the data marts and charged them with modeling the enterprise-wide database that would be the foundation of the data warehouse. They had to make sure that the data would always be up to date, consistent for the entire enterprise, and complete.

The company launched its enterprise data warehouse in May 2006 to coincide with its consolidation of applications and data centers. To date, HP has consolidated hundreds of data marts into just over 200. The data warehouse contains 180 terabytes of raw data and 75 terabytes of functional data. At some point in 2008, the size should double and the data warehouse will be complete. Fifty thousand HP workers will utilize the data warehouse. All HP financial data will be able to be accessed via the data warehouse.

HP believes so strongly in its development of the system that the company is trying to sell its expertise to other companies that are seeking data warehouse technology. HP is marketing a product called Neoview, which has been developed from the proprietary work that the company has done in creating its own data warehouse.

More than 100 database specialists and software developers at HP are perfecting the system’s dexterity with table joins and giving it the ability to perform analysis functions at the same time that it is managing new incoming data. HP is also enhancing Neoview’s management and monitoring tools.

The servers employed by the Neoview system utilize Itanium processors from Intel, so they meet industry standards and are far more versatile than servers with proprietary technology. The system is also highly scalable and promises availability 99.999 percent of the time.

HP’s first customer for the Neoview database system was Bon-Ton Stores, which operates 272 department stores and 7 furniture stores in 23 states. Bon-Ton purchased a 7-terabyte Neoview system for a data warehouse that includes merchandise, customer, and supplier data for merchandise analysis and marketing. One of the database tables for the warehouse has more than 4 billion rows. Bon-Ton’s CIO James Lance reported that the Neoview system exceeded expectations.

Wal-Mart signed up for Neoview to work with its strategic Retail Link system, which allows its 20,000 suppliers to access data about the movement and sales of their products in its stores. For over a decade Wal-Mart has operated one of the largest commercial data warehouses in the world with more than 1,000 Terabytes of sales information on every item sold in its stores.
Wal-Mart uses the data warehouse to analyze in-store sales, but it would like to do more with the data to determine the ideal mix of items for each store’s customers and to place these items in stores where they are most likely to be purchased.

HP has been able to sell the Neoview technology by differentiating it from typical data warehouses, which are costly, use proprietary technology, and tend to focus on one area of a business rather than an entire enterprise. For example, airlines have data warehouses for yield management and telecommunications carriers have data warehouses to minimize customer attrition. A true enterprise data warehouse would have all these entities plus data on employees, customer service, marketing campaigns, and financial reporting—in other words, all of the data used by the company.

Wal-Mart had been using Teradata’s data warehousing platform to support Retail Link. It will continue using Teradata, but will allow Neoview to shoulder some of the workload. Wal-Mart’s chief technology officer (CTO) Nancy Stewart reported that selection of Neoview was a “price-performance decision.” After several months of testing production loads and accuracy of query results, Stewart reported that “Neoview fits right into that environment of extreme high availability and high performance.” Wal-Mart put Neoview into production in early June 2007.

Very few companies have built an all-inclusive data warehouse serving the entire enterprise. They require enormous work to organize and integrate all the data as well as knowledge of database technology and design principles. Businesses are changing constantly due to corporate mergers and global expansion, making today’s data warehouse out of date tomorrow. There are also political turf issues. Not all departments want to depend on a central data warehouse supported by a centralized information systems staff for their data-analysis needs. All of HP’s departmental users initially resisted the idea of a central data warehouse.

HP will emphasize cost and flexibility. Neoview’s hardware can be used to run other applications aside from those connected to the data warehouse. The company does not see a current product that serves as the comprehensive enterprise data warehouse that it intends Neoview to be. Other data warehouses in use do not come close to incorporating 100 percent of a company’s data, which is HP’s goal for its own data warehouse and Neoview.

HP anticipates luring companies to Neoview that have not been previously interested in such technology by beating the competition on price and simplicity. Using industry-standard hardware should complement this strategy. Most information technology managers should be able to work with the familiar components of Neoview, whereas the pool of people with the knowledge to run most data warehouses is quite shallow. According to database expert Jim Gray, “Right now, it takes far too much expertise to install and use the systems for data mining.”

One way in which HP intends to address the issue of complexity is to preconfigure Neoview installations for a particular industry and the workload and applications that the industry requires. Neoview will be a complete solution in a box with on-site management and remote support from HP. Mott hopes to convince potential customers that the cost of running multiple data marts is equal to or greater than the cost of a data warehouse, which will ultimately be a better solution for a business.


Case Study Questions
1. Identify the problem described in this case. What people, organization, and technology factors were responsible for creating this problem?
2. What solution has HP chosen to fix this problem? Did management select the best solution alternative?
3. How much will HP’s database experience and technology help HP and its clients build all-inclusive data warehouses?
5. If you were in charge of developing an enterprise-wide data warehouse for your company, describe the steps you would have to take to complete this project. List and describe all people, organization, and technology issues that must be addressed to build an enterprise-wide data warehouse successfully.