Chapter 6

Foundations of Business Intelligence: Databases and Information Management

LEARNING OBJECTIVES

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

1. What are the problems of managing data resources in a traditional file environment and how are they solved by a database management system?
2. What are the major capabilities of database management systems (DBMS) and why is a relational DBMS so powerful?
3. What are some important principles of database design?
4. What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?
5. Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?

Interactive Sessions:

Big Data, Big Rewards

Controversy Whirls Around the Consumer Product Safety Database

CHAPTER OUTLINE

BAE SYSTEMS

BAE Systems (BAE) is the United Kingdom's largest manufacturing company and one of the largest commercial aerospace and defence organisations in Europe. Its high-technology, information-driven products and services range from one of the world's most capable multi-role combat fighters, the Eurofighter Typhoon, to the Jetstream family of commercial aircraft, to the provision of information technology and information systems for e-business to develop and implement logistics, IT and e-capability services. With sales, manufacturing and support sites throughout the world, including the U.K., Europe, the United States, and Australia, BAE employs 88,000 people and generates more than U.S. $ 30 billion in annual revenue.

Although BAE has consolidated its competitive position in established markets, and continues to expand into new markets in the Middle East and Asia, its performance in the aircraft part of the business was being impeded by legacy information systems which support the computer-aided design (CAD) and computer-aided manufacturing (CAM) of its aircraft. The distributed nature of BAE's design and manufacturing sites meant that storing and analysing accurate sets of operational data describing the complex components of the various aircraft types to produce aircraft assembly reports for the production lines became increasingly challenging and resource-consuming. Data describing the same aircraft component parts might need resolution, such as in the case of various part naming conventions and codes.

Accessing the data from the many systems was a complex task involving many technical challenges. As the aircraft business of BAE grew so did the likelihood for delays in producing the aircraft assembly reports and other operations data sets necessary for aircraft production management decision making. In the worst case, the production of aircraft on the assembly line would stop until accurate information was available, with consequent schedule and cost implications. BAE's CAD/CAM staff were storing and analysing data sets sourced from 5 major aircraft design and manufacturing sites spread throughout the U.K., each host to thousands of staff involved in the design and manufacturing process, so that assembly reports and other operations data could be produced. Although the data that the legacy systems processed were held principally in computer files, there were numerous occasions when paper drawings with annotations containing component design and manufacturing information were used to reconcile ambiguities and inconsistencies in the assembly reports. When these data ambiguities and inconsistencies occurred, this gave rise to a sense of uncertainty in the assembly reports produced.

© Kristoffer Tripplaar/Alamy

What BAE needed was a single repository for CAD/CAM data that would also facilitate the integration of data held in its legacy systems. The company decided to replace its legacy systems with an enterprise-wide knowledge management system which would bring the design and manu-

facturing data into a single database that could be concurrently accessed by the design and manufacturing engineers. BAE implemented Siemens' Teamcenter product lifecycle management software and Dassault Systemes' CATIA CAD/CAM software. Teamcenter can also be configured to take advantage of recent developments in cloud computing using Microsoft's Azure, IBM's SmartCloud Enterprise + , and Amazon Web Services.

Bringing together Siemens' Teamcenter and Dassault Systemes' CATIA has given BAE Systems powerful integrated data management tools. The Teamcenter database includes tools for component markup and rollup capabilities allowing users to visualise the effect of component design changes and configuration selections in real-time.

The new solution has produced significant cost savings at BAE in terms of its design and manufacturing data management and storage, while boosting performance. With fewer legacy systems and data files to manage, BAE has been able to meet quality, time and cost requirements by being able to produce complete and accurate aircraft component definitions and configurations. BAE's new design and manufacturing database technology has improved speed-to-market by synchronising upstream CAD and downstream CAM component definitions, thereby enabling better cross-discipline coordination. With these savings, the company has been able to spend more resources on improving data management across the entire enterprise.

Sources: "BAE Systems Half-Yearly Report and Presentation 2012" www.baesystems.com, accessed November 8, 2012; "Teamcenter supports aircraft through 50-year cycle: BAE Systems Military Air Solutions" www.plm.automation.siemens.com, accessed November 8, 2012; "CATIA V5 Fact Sheet" www.3ds.com, accessed November 8, 2012.

Case contributed by Robert Manderson, University of Roehampton

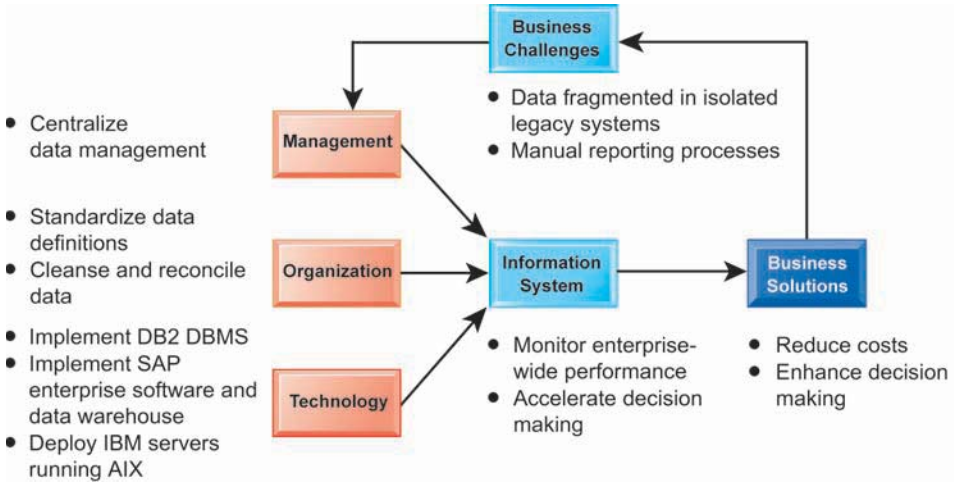
The experience of BAE Systems illustrates the importance of data management. Business performance depends on the accuracy and reliability of its data. The company has grown its business, but, both operational CAD/CAM efficiency and production management decision making were impeded by data stored in legacy systems that were difficult to access. How businesses store, organise, and manage their data has a huge impact on organisational effectiveness.

The chapter-opening diagram calls attention to important points raised by this case and this chapter. BAE Systems management decided that the firm needed to improve the management of its data. Pieces of data about design components, manufactured components, and their final assembly had been stored in many large legacy systems that made it extremely difficult for the data to be retrieved, correctly unified so that it could be used in the production line assembly of aircraft components. The data were often redundant and inconsistent, limiting their usefulness. Management was unable to obtain an enterprise-view of the company.

In the past, BAE Systems had used manual paper processes to reconcile its inconsistent and redundant data and to assemble data for management reporting. This solution was extremely time-consuming and costly and prevented the company's information technology department from performing higher-value work. A more appropriate solution was to install new hardware and software to create an enterprise-wide repository for business information that would support a more streamlined set of business applications. The new software included enterprise software that was integrated with an up-to-date database management system that could supply data for enterprise-wide reporting. The company had to reorganise its data into a standard company-wide format, eliminate redundancies, and establish rules, responsibilities, and procedures for updating and using the data.

A state-of-the-art database management system suite of software helps BAE Systems boost efficiency by making it easier to locate and assemble data for management reporting and for processing day-to-day CAD/CAM transactions for final aircraft component assembly. The data are more accurate and reliable, and costs for managing and storing the data have been considerably reduced.

Here are some questions to think about: What kinds of data management problems did BAE Systems experience in its legacy database environment? What work had to be done before the company could effectively take advantage of the new data management technology?



6.1 Organizing Data in a Traditional File Environment

An effective information system provides users with accurate, timely, and relevant information. Accurate information is free of errors. Information is timely when it is available to decision makers when it is needed. Information is relevant when it is useful and appropriate for the types of work and decisions that require it.

You might be surprised to learn that many businesses don't have timely, accurate, or relevant information because the data in their information systems have been poorly organized and maintained. That's why data management is so essential. To understand the problem, let's look at how information systems arrange data in computer files and traditional methods of file management.

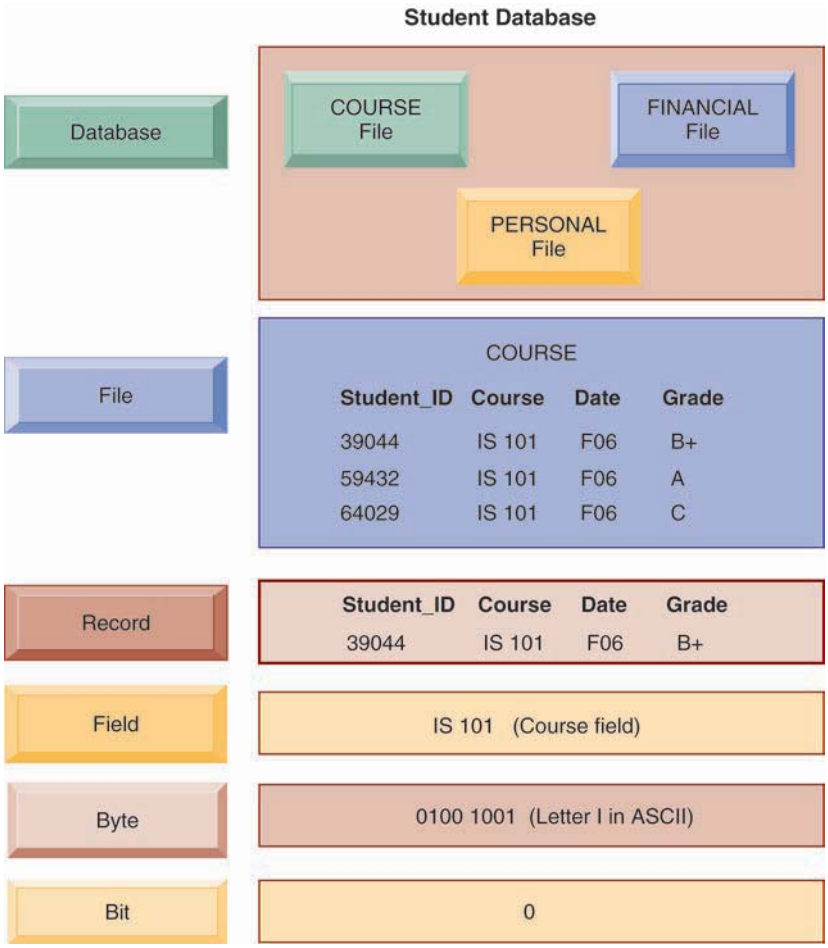
FILE ORGANIZATION TERMS AND CONCEPTS

A computer system organizes data in a hierarchy that starts with bits and bytes and progresses to fields, records, files, and databases (see Figure 6.1). A **bit** represents the smallest unit of data a computer can handle. A group of bits, called a **byte**, represents a single character, which can be a letter, a number, or another symbol. A grouping of characters into a word, a group of words, or a complete number (such as a person's name or age) is called a **field**. A group of related fields, such as the student's name, the course taken, the date, and the grade, comprises a **record**; a group of records of the same type is called a **file**.

For example, the records in Figure 6.1 could constitute a student course file. A group of related files makes up a database. The student course file illustrated in Figure 6.1 could be grouped with files on students' personal histories and financial backgrounds to create a student database.

A record describes an entity. An **entity** is a person, place, thing, or event on which we store and maintain information. Each characteristic or quality describing a particular entity is called an **attribute**. For example, Student_ID, Course, Date, and Grade are attributes of the entity COURSE. The specific values that these attributes can have are found in the fields of the record describing the entity COURSE.

FIGURE 6.1 THE DATA HIERARCHY



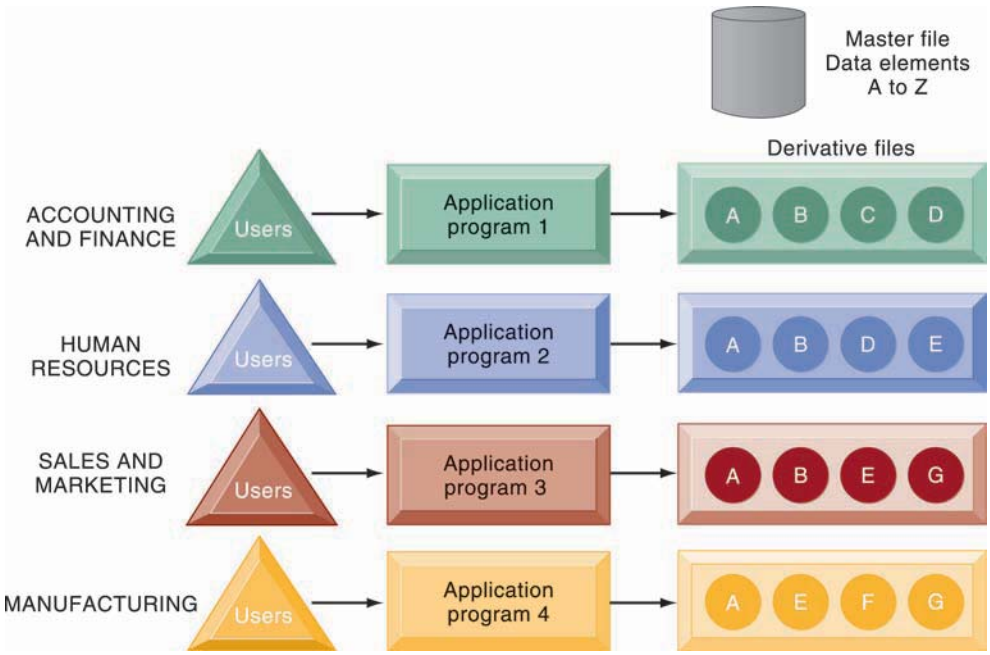
A computer system organizes data in a hierarchy that starts with the bit, which represents either a 0 or a 1. Bits can be grouped to form a byte to represent one character, number, or symbol. Bytes can be grouped to form a field, and related fields can be grouped to form a record. Related records can be collected to form a file, and related files can be organized into a database.

PROBLEMS WITH THE TRADITIONAL FILE ENVIRONMENT

In most organizations, systems tended to grow independently without a company-wide plan. Accounting, finance, manufacturing, human resources, and sales and marketing all developed their own systems and data files. Figure 6.2 illustrates the traditional approach to information processing.

Each application, of course, required its own files and its own computer program to operate. For example, the human resources functional area might have a personnel master file, a payroll file, a medical insurance file, a pension file, a mailing list file, and so forth until tens, perhaps hundreds, of files and programs existed. In the company as a whole, this process led to multiple master files created, maintained, and operated by separate divisions or departments. As this process goes on for 5 or 10 years, the organization is saddled with hundreds of programs and applications that are very difficult to maintain and manage. The resulting problems are data redundancy and inconsistency,

FIGURE 6.2 TRADITIONAL FILE PROCESSING



The use of a traditional approach to file processing encourages each functional area in a corporation to develop specialized applications. Each application requires a unique data file that is likely to be a subset of the master file. These subsets of the master file lead to data redundancy and inconsistency, processing inflexibility, and wasted storage resources.

program-data dependence, inflexibility, poor data security, and an inability to share data among applications.

Data Redundancy and Inconsistency

Data redundancy is the presence of duplicate data in multiple data files so that the same data are stored in more than one place or location. Data redundancy occurs when different groups in an organization independently collect the same piece of data and store it independently of each other. Data redundancy wastes storage resources and also leads to **data inconsistency**, where the same attribute may have different values. For example, in instances of the entity COURSE illustrated in Figure 6.1, the Date may be updated in some systems but not in others. The same attribute, Student_ID, may also have different names in different systems throughout the organization. Some systems might use Student_ID and others might use ID, for example.

Additional confusion might result from using different coding systems to represent values for an attribute. For instance, the sales, inventory, and manufacturing systems of a clothing retailer might use different codes to represent clothing size. One system might represent clothing size as "extra large," whereas another might use the code "XL" for the same purpose. The resulting confusion would make it difficult for companies to create customer relationship management, supply chain management, or enterprise systems that integrate data from different sources.

Program-Data Dependence

Program-data dependence refers to the coupling of data stored in files and the specific programs required to update and maintain those files such that changes in programs require changes to the data. Every traditional computer program has to describe the location and nature of the data with which it works. In a traditional file environment, any change in a software program could require a change in the data accessed by that program. One program might be modified from a five-digit to a nine-digit zip code. If the original data file were changed from five-digit to nine-digit zip codes, then other programs that required the five-digit zip code would no longer work properly. Such changes could cost millions of dollars to implement properly.

Lack of Flexibility

A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad hoc reports or respond to unanticipated information requirements in a timely fashion. The information required by ad hoc requests is somewhere in the system but may be too expensive to retrieve. Several programmers might have to work for weeks to put together the required data items in a new file.

Poor Security

Because there is little control or management of data, access to and dissemination of information may be out of control. Management may have no way of knowing who is accessing or even making changes to the organization's data.

Lack of Data Sharing and Availability

Because pieces of information in different files and different parts of the organization cannot be related to one another, it is virtually impossible for information to be shared or accessed in a timely manner. Information cannot flow freely across different functional areas or different parts of the organization. If users find different values of the same piece of information in two different systems, they may not want to use these systems because they cannot trust the accuracy of their data.

6.2 The Database Approach to Data Management

Database technology cuts through many of the problems of traditional file organization. A more rigorous definition of a **database** is a collection of data organized to serve many applications efficiently by centralizing the data and controlling redundant data. Rather than storing data in separate files for each application, data appears to users as being stored in only one location. A single database services multiple applications. For example, instead of a corporation storing employee data in separate information systems and separate files for personnel, payroll, and benefits, the corporation could create a single common human resources database.

DATABASE MANAGEMENT SYSTEMS

A **database management system (DBMS)** is software that permits an organization to centralize data, manage them efficiently, and provide access

to the stored data by application programs. The DBMS acts as an interface between application programs and the physical data files. When the application program calls for a data item, such as gross pay, the DBMS finds this item in the database and presents it to the application program. Using traditional data files, the programmer would have to specify the size and format of each data element used in the program and then tell the computer where they were located.

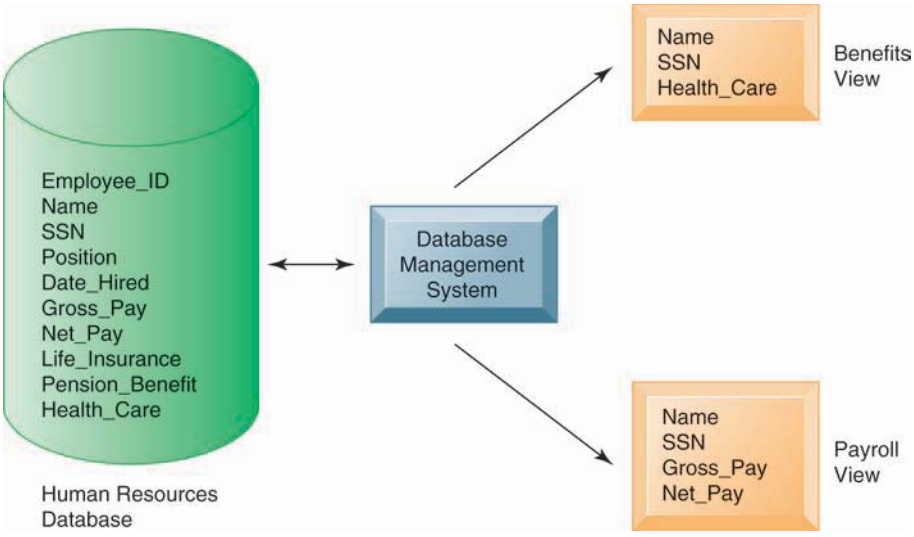
The DBMS relieves the programmer or end user from the task of understanding where and how the data are actually stored by separating the logical and physical views of the data. The *logical view* presents data as they would be perceived by end users or business specialists, whereas the *physical view*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

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 6.3, 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 these views are stored in a single database, where they can be more easily managed by the organization.

How a DBMS Solves the Problems of the Traditional File Environment

A DBMS reduces data redundancy and inconsistency by minimizing isolated files in which the same data are repeated. The DBMS may not enable the organization to eliminate data redundancy entirely, but it can help control redundancy. Even if the organization maintains some redundant data, using a DBMS eliminates data inconsistency because the DBMS can help the organization ensure that every occurrence of redundant data has the same values. The DBMS uncouples programs and data, enabling data to stand

FIGURE 6.3 HUMAN RESOURCES DATABASE WITH MULTIPLE VIEWS



A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

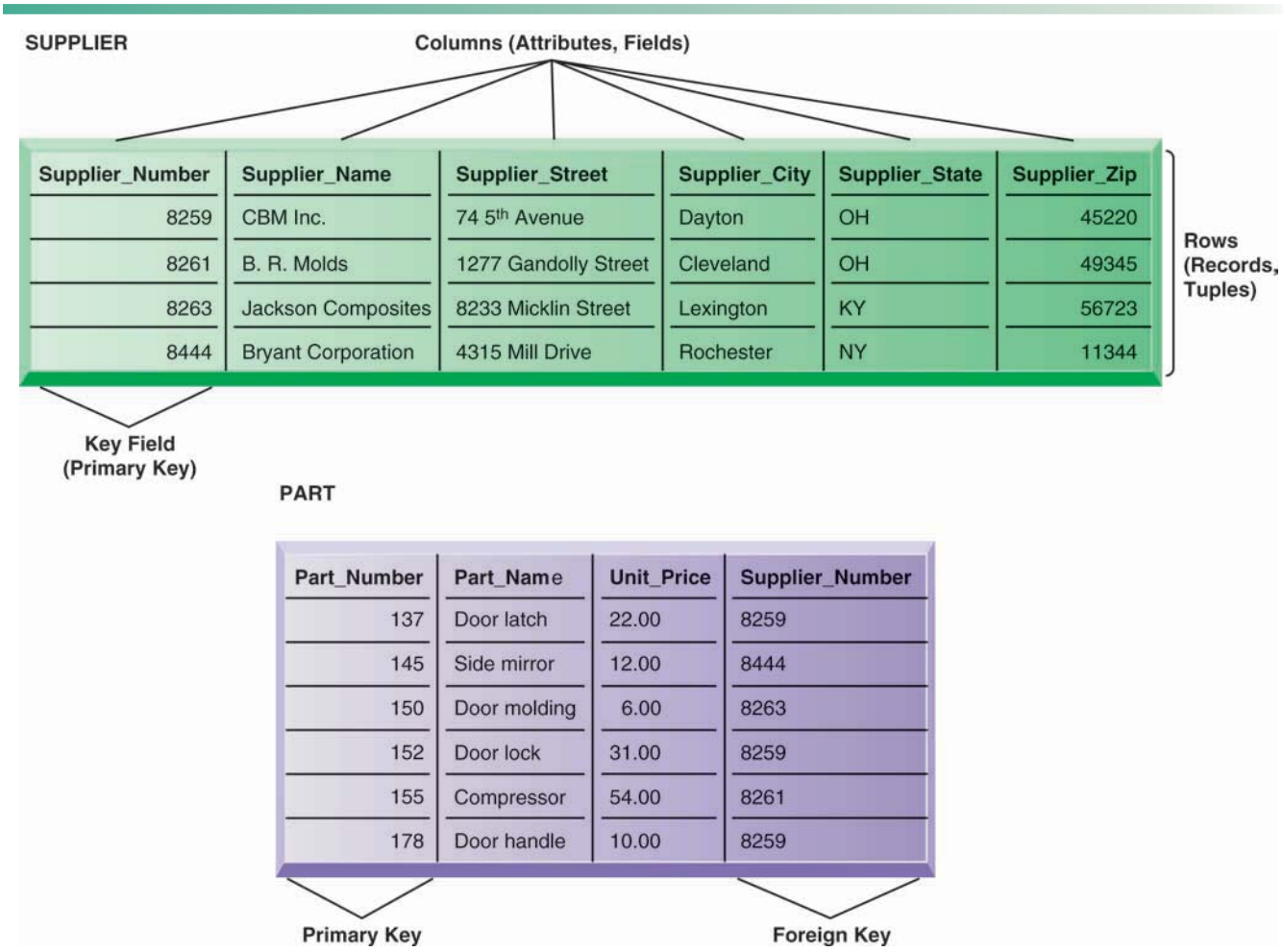
on their own. Access and availability of information will be increased and program development and maintenance costs reduced because users and programmers can perform ad hoc queries of data in the database. The DBMS enables the organization to centrally manage data, their use, and security.

Relational DBMS

Contemporary DBMS use different database models to keep track of entities, attributes, and relationships. The most popular type of DBMS today for PCs as well as for larger computers and mainframes is the **relational DBMS**. Relational databases represent data as two-dimensional tables (called relations). Tables may be referred to as files. Each table contains data on an entity and its attributes. Microsoft Access is a relational DBMS for desktop systems, whereas DB2, Oracle Database, and Microsoft SQL Server are relational DBMS for large mainframes and midrange computers. MySQL is a popular open source DBMS, and Oracle Database Lite is a DBMS for mobile computing devices.

Let's look at how a relational database organizes data about suppliers and parts (see Figure 6.4). The database has a separate table for the entity SUPPLIER and a table for the entity PART. Each table consists of a grid of columns and

FIGURE 6.4 RELATIONAL DATABASE TABLES



A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier_Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

rows of data. Each individual element of data for each entity is stored as a separate field, and each field represents an attribute for that entity. Fields in a relational database are also called columns. For the entity SUPPLIER, the supplier identification number, name, street, city, state, and zip code are stored as separate fields within the SUPPLIER table and each field represents an attribute for the entity SUPPLIER.

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**. Data for the entity PART have their own separate table.

The field for Supplier_Number in the SUPPLIER table uniquely identifies each record so that the record can be retrieved, updated, or sorted. 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. Supplier_Number is the primary key for the SUPPLIER table and Part_Number is the primary key for the PART table. Note 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 lookup field to look up data about the supplier of a specific part.

Operations of a Relational DBMS

Relational database tables can be combined easily to deliver data required by users, provided that any two tables share a common data element. 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 files have a shared data element: Supplier_Number.

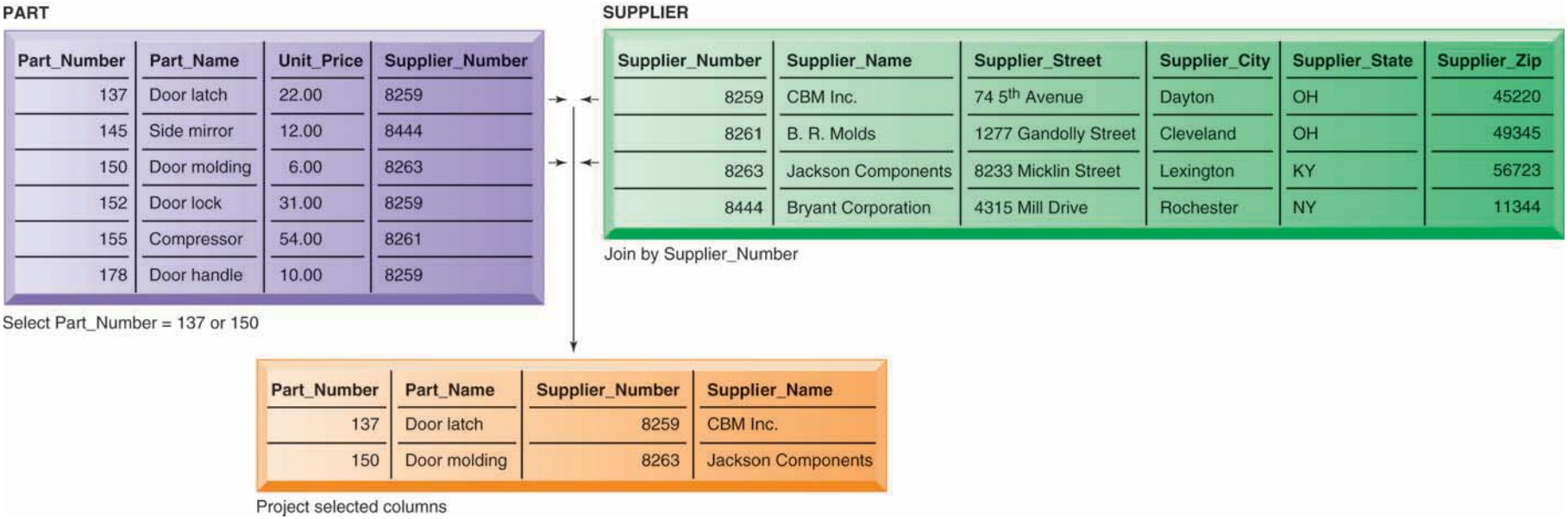
In a relational database, three basic operations, as shown in Figure 6.5, are used to develop useful sets of data: select, join, and project. The *select* operation creates a subset consisting of all records in the file that meet stated criteria. Select creates, in other words, a subset of rows that 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.

Non-Relational Databases and Databases in the Cloud

For over 30 years, relational database technology has been the gold standard. Cloud computing, unprecedented data volumes, massive workloads for Web services, and the need to store new types of data require database alternatives to the traditional relational model of organizing data in the form of tables, columns, and rows. Companies are turning to "*NoSQL*" non-relational database technologies for this purpose. **Non-relational database management systems** use a more flexible data model and are designed for managing large data sets across many distributed machines and for easily scaling up or down. They are useful for accelerating simple queries against large volumes of structured and

FIGURE 6.5 THE THREE BASIC OPERATIONS OF A RELATIONAL DBMS



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

unstructured data, including Web, social media, graphics, and other forms of data that are difficult to analyze with traditional SQL-based tools.

There are several different kinds of NoSQL databases, each with its own technical features and behavior. Oracle NoSQL Database is one example, as is Amazon's SimpleDB, one of the Amazon Web Services that run in the cloud. SimpleDB provides a simple Web services interface to create and store multiple data sets, query data easily, and return the results. There is no need to pre-define a formal database structure or change that definition if new data are added later.

Amazon and other cloud computing vendors provide relational database services as well. Amazon Relational Database Service (Amazon RDS) offers MySQL, SQL Server, or Oracle Database as database engines. Pricing is based on usage. Oracle has its own Database Cloud Service using its relational Oracle Database 11g, and Microsoft SQL Azure Database is a cloud-based relational database service based on Microsoft's SQL Server DBMS. Cloud-based data management services have special appeal for Web-focused start-ups or small to medium-sized businesses seeking database capabilities at a lower price than in-house database products.

TicketDirect, which sells tickets to concerts, sporting events, theater performances, and movies in Australia and New Zealand, adopted the SQL Azure Database cloud platform in order to improve management of peak system loads during major ticket sales. It migrated its data to the SQL Azure database. By moving to a cloud solution, TicketDirect is able to scale its computing resources in response to real-time demand while keeping costs low.

In addition to public cloud-based data management services, companies now have the option of using databases in private clouds. For example, Sabre Holdings, the world's largest software as a service (SaaS) provider for the aviation industry, has a private database cloud that supports more than 100 projects and 700 users. A consolidated database spanning a pool of standardized servers running Oracle Database 11g provides database services for multiple applications. Workload management tools ensure sufficient resources are available to meet application needs even when the workload changes. The shared hardware and software platform reduces the number of servers, DBMS, and storage devices needed for these projects, which consist of custom airline travel applications along with rail, hotel, and other travel industry applications (Baum, 2011).

Private clouds consolidate servers, storage, operating systems, databases, and mixed workloads onto a shared hardware and software infrastructure. Deploying databases on a consolidated private cloud enables IT departments to improve quality of service levels and reduce capital and operating costs. The higher the consolidation density achieved, the greater the return on investment.

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 language, 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 6.6). Data dictionaries for large corporate databases may capture additional information, such as usage, ownership (who in the organization is responsible for maintaining the data), authorization, security, and the individuals, business functions, programs, and reports that use each data element.

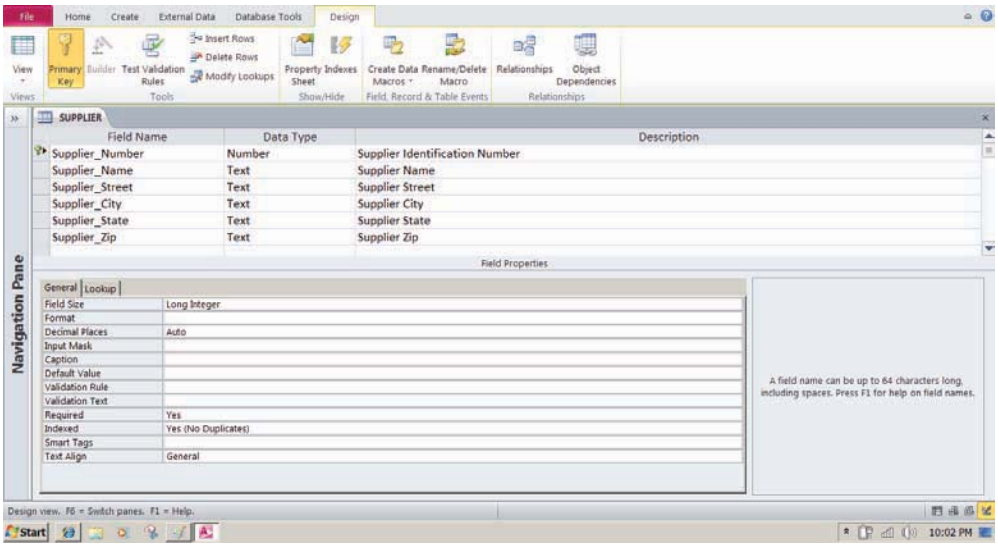
Querying and Reporting

DBMS includes 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 6.7 illustrates the SQL query that would produce the new resultant table in Figure 6.5. You can find out more about how to perform SQL queries in our Learning Tracks for this chapter.

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.

In Microsoft Access, you will find features that enable users to create queries by identifying the tables and fields they want and the results, and then selecting the rows from the database that meet particular criteria. These actions in turn are translated into SQL commands. Figure 6.8 illustrates how the same query as the SQL query to select parts and suppliers would be constructed using the Microsoft query-building tools.

FIGURE 6.6 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 6.7 EXAMPLE OF AN SQL QUERY

```
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;
```

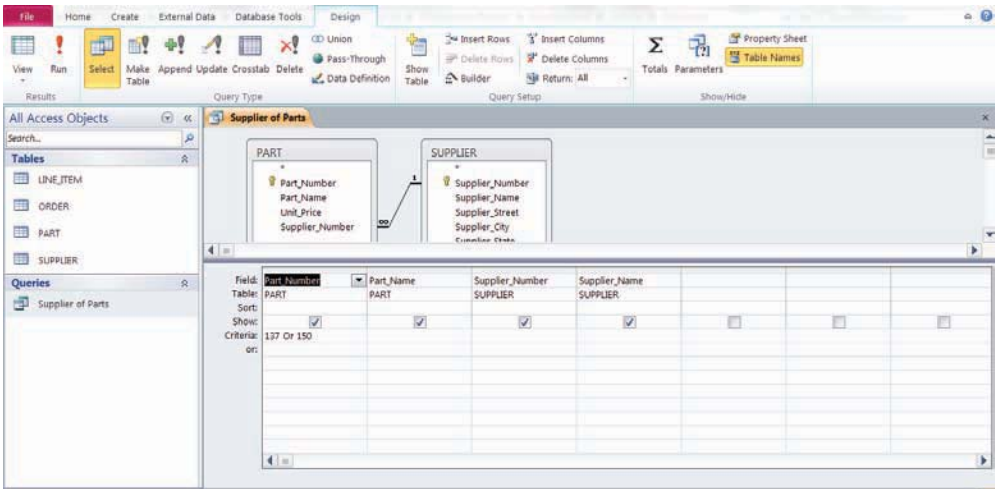
Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a list with the same results as Figure 6.5.

Microsoft Access and other DBMS 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 Access. 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.

DESIGNING DATABASES

To create a database, you must understand the relationships among the data, the type of data that will be maintained in the database, how the data will be used, and how the organization will need to change to manage data from a company-wide perspective. The database requires both a conceptual design and a physical design. The conceptual, or logical, design of a database is an abstract model of the database from a business perspective, whereas the physical design shows how the database is actually arranged on direct-access storage devices.

FIGURE 6.8 AN ACCESS QUERY



Illustrated here is how the query in Figure 6.7 would be constructed using Microsoft Access query-building tools. It shows the tables, fields, and selection criteria used for the query.

FIGURE 6.9 AN UNNORMALIZED RELATION FOR ORDER

An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

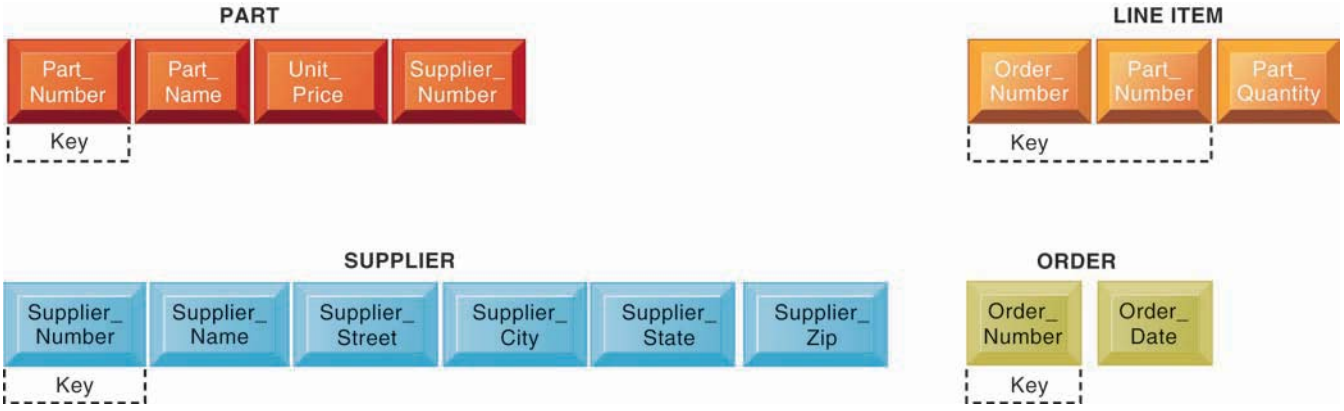
Normalization and Entity-Relationship Diagrams

The conceptual database design describes how the data elements in the database are to be grouped. The design process identifies relationships among data elements and the most efficient way of grouping data elements together to meet business information requirements. The process also identifies redundant data elements and the groupings of data elements required for specific application programs. Groups of data are organized, refined, and streamlined until an overall logical view of the relationships among all the data in the database emerges.

To use a relational database model effectively, complex groupings of data must be streamlined to minimize redundant data elements and awkward many-to-many relationships. The process of creating small, stable, yet flexible and adaptive data structures from complex groups of data is called **normalization**. Figures 6.9 and 6.10 illustrate this process.

In the particular business modeled here, an order can have more than one part but each part is provided by only one supplier. If we build a relation called ORDER with all the fields included here, we would have to repeat the name and address of the supplier for every part on the order, even though the order is for parts from a single supplier. This relationship contains what are called repeating data groups because there can be many parts on a single order to a given supplier. A more efficient way to arrange the data is to break down ORDER into smaller relations, each of which describes a single entity. If we go step by step and normalize the relation ORDER, we emerge with the relations illustrated in Figure 6.10. You can find out more about normalization,

FIGURE 6.10 NORMALIZED TABLES CREATED FROM ORDER



After normalization, the original relation ORDER has been broken down into four smaller relations. The relation ORDER is left with only two attributes and the relation LINE_ITEM has a combined, or concatenated, key consisting of Order_Number and Part_Number.

entity-relationship diagramming, and database design in the Learning Tracks for this chapter.

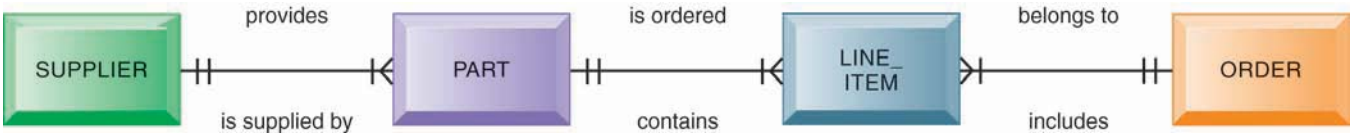
Relational database systems try to enforce **referential integrity** rules to ensure that relationships between coupled tables remain consistent. When one table has a foreign key that points to another table, you may not add a record to the table with the foreign key unless there is a corresponding record in the linked table. In the database we examined earlier in this chapter, 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!

Database designers document their data model with an **entity-relationship diagram**, illustrated in Figure 6.11. This diagram illustrates the relationship between the entities SUPPLIER, PART, LINE_ITEM, and ORDER. 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 6.11 shows that one ORDER can contain many LINE_ITEMs. (A PART can be ordered many times and appear many times as a line item in a single order.) Each PART can have only one SUPPLIER, but many PARTs can be provided by the same SUPPLIER.

It can't be emphasized enough: If the business doesn't get its data model right, the system won't be able to serve the business well. The company's systems will not be as effective as they could be because they'll 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 a 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.

FIGURE 6.11 AN ENTITY-RELATIONSHIP DIAGRAM



This diagram shows the relationships between the entities SUPPLIER, PART, LINE_ITEM, and ORDER that might be used to model the database in Figure 6.10.

6.3 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, keeping track of 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.

THE CHALLENGE OF BIG DATA

Up until about five years ago, most data collected by organizations consisted of transaction data that could easily fit into rows and columns of relational database management systems. Since then, there has been an explosion of data from Web traffic, e-mail messages, and social media content (tweets, status messages), as well as machine-generated data from sensors (used in smart meters, manufacturing sensors, and electrical meters) or from electronic trading systems. These data may be unstructured or semi-structured and thus not suitable for relational database products that organize data in the form of columns and rows. We now use the term **big data** to describe these datasets with volumes so huge that they are beyond the ability of typical DBMS to capture, store, and analyze.

Big data doesn't refer to any specific quantity, but usually refers to data in the petabyte and exabyte range—in other words, billions to trillions of records, all from different sources. Big data are produced in much larger quantities and much more rapidly than traditional data. For example, a single jet engine is capable of generating 10 terabytes of data in just 30 minutes, and there are more than 25,000 airline flights each day. Even though "tweets" are limited to 140 characters each, Twitter generates over 8 terabytes of data daily. According to the International Data Center (IDC) technology research firm, data are more than doubling every two years, so the amount of data available to organizations is skyrocketing.

Businesses are interested in big data because they can reveal more patterns and interesting anomalies than smaller data sets, with the potential to provide new insights into customer behavior, weather patterns, financial market activity, or other phenomena. However, to derive business value from these data, organizations need new technologies and tools capable of managing and analyzing non-traditional data along with their traditional enterprise data.

BUSINESS INTELLIGENCE INFRASTRUCTURE

Suppose you wanted concise, reliable information about current operations, trends, and changes across the entire company. If you worked in a large company, the data you need might have to be pieced together from separate systems, such as sales, manufacturing, and accounting, and even from external sources, such as demographic or competitor data. Increasingly, you might need to use big data. A contemporary infrastructure for business intelligence has an array of tools for obtaining useful information from all the different types of data used by businesses today, including semi-structured and unstructured big data in vast quantities. These capabilities include data warehouses and data marts, Hadoop, in-memory computing, and analytical platforms.

Data Warehouses and Data Marts

The traditional tool for analyzing corporate data for the past two decades has been the 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 extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and transformed by correcting inaccurate and incomplete data and restructuring the data for management reporting and analysis before being loaded into the data warehouse.

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 .

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. Bookseller Barnes & Noble used to maintain a series of data marts—one for point-of-sale data in retail stores, another for college bookstore sales, and a third for online sales.

Hadoop

Relational DBMS and data warehouse products are not well-suited for organizing and analyzing big data or data that do not easily fit into columns and rows used in their data models. For handling unstructured and semi-structured data in vast quantities, as well as structured data, organizations are using **Hadoop**. Hadoop is an open source software framework managed by the Apache Software Foundation that enables distributed parallel processing of huge amounts of data across inexpensive computers. It breaks a big data problem down into sub-problems, distributes them among up to thousands of inexpensive computer processing nodes, and then combines the result into a smaller data set that is easier to analyze. You've probably used Hadoop to find the best airfare on the Internet, get directions to a restaurant, do a search on Google, or connect with a friend on Facebook.

Hadoop consists of several key services: the Hadoop Distributed File System (HDFS) for data storage and MapReduce for high-performance parallel data processing. HDFS links together the file systems on the numerous nodes in a Hadoop cluster to turn them into one big file system. Hadoop's MapReduce was inspired by Google's MapReduce system for breaking down processing of huge datasets and assigning work to the various nodes in a cluster. HBase, Hadoop's non-relational database, provides rapid access to the data stored on HDFS and a transactional platform for running high-scale real-time applications.

Hadoop can process large quantities of any kind of data, including structured transactional data, loosely structured data such as Facebook and Twitter feeds, complex data such as Web server log files, and unstructured audio and video data. Hadoop runs on a cluster of inexpensive servers, and processors can be added or removed as needed. Companies use Hadoop for analyzing very large

volumes of data as well as for a staging area for unstructured and semi-structured data before they are loaded into a data warehouse. Facebook stores much of its data on its massive Hadoop cluster, which holds an estimated 100 petabytes, about 10,000 times more information than the Library of Congress. Yahoo uses Hadoop to track user behavior so it can modify its home page to fit their interests. Life sciences research firm NextBio uses Hadoop and HBase to process data for pharmaceutical companies conducting genomic research. Top database vendors such as IBM, Hewlett-Packard, Oracle, and Microsoft have their own Hadoop software distributions. Other vendors offer tools for moving data into and out of Hadoop or for analyzing data within Hadoop.

In-Memory Computing

Another way of facilitating big data analysis is to use **in-memory computing**, which relies primarily on a computer's main memory (RAM) for data storage. (Conventional DBMS use disk storage systems.) Users access data stored in system primary memory, thereby eliminating bottlenecks from retrieving and reading data in a traditional, disk-based database and dramatically shortening query response times. In-memory processing makes it possible for very large sets of data, amounting to the size of a data mart or small data warehouse, to reside entirely in memory. Complex business calculations that used to take hours or days are able to be completed within seconds, and this can even be accomplished on handheld devices.

The previous chapter describes some of the advances in contemporary computer hardware technology that make in-memory processing possible, such as powerful high-speed processors, multicore processing, and falling computer memory prices. These technologies help companies optimize the use of memory and accelerate processing performance while lowering costs.

Leading commercial products for in-memory computing include SAP's High Performance Analytics Appliance (HANA) and Oracle Exalytics. Each provides a set of integrated software components, including in-memory database software and specialized analytics software, that run on hardware optimized for in-memory computing work.

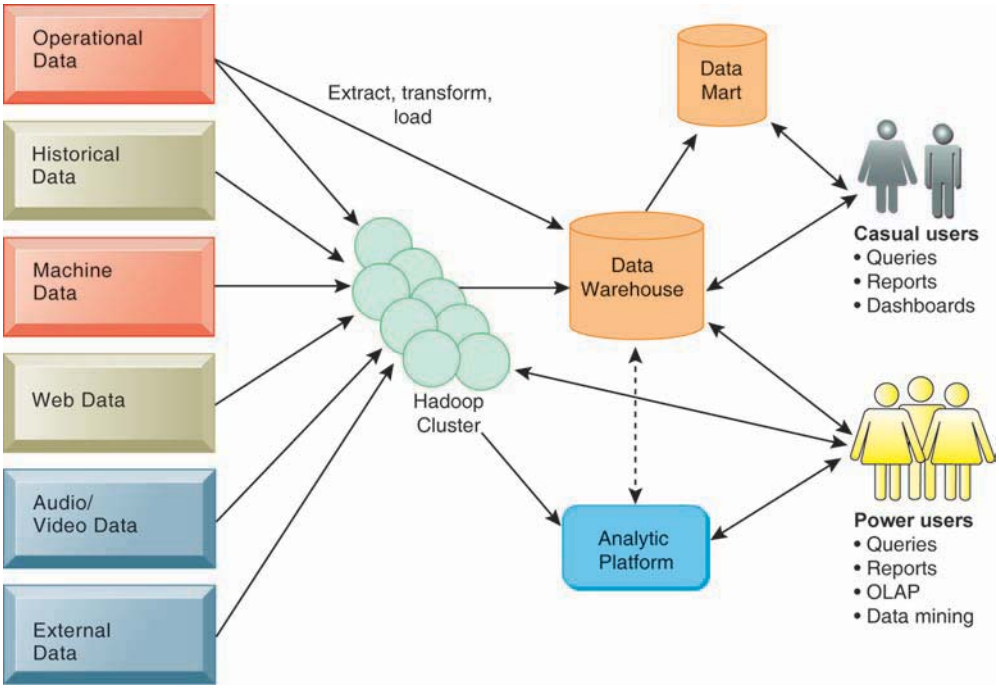
Centrica, a gas and electric utility, uses HANA to quickly capture and analyze the vast amounts of data generated by smart meters. The company is able to analyze usage every 15 minutes, giving it a much clearer picture of usage by neighborhood, home size, type of business served, or building type. HANA also helps Centrica show its customers their energy usage patterns in real-time using online and mobile tools.

Analytic Platforms

Commercial database vendors have developed specialized high-speed **analytic platforms** using both relational and non-relational technology that are optimized for analyzing large datasets. These analytic platforms, such as IBM Netezza and Oracle Exadata, feature preconfigured hardware-software systems that are specifically designed for query processing and analytics. For example, IBM Netezza features tightly integrated database, server, and storage components that handle complex analytic queries 10 to 100 times faster than traditional systems. Analytic platforms also include in-memory systems and NoSQL non-relational database management systems.

Figure 6.12 illustrates a contemporary business intelligence infrastructure using the technologies we have just described. Current and historical data are extracted from multiple operational systems along with Web data, machine-generated data, unstructured audio/visual data, and data from external sources

FIGURE 6.12 COMPONENTS OF A DATA WAREHOUSE



A contemporary business intelligence infrastructure features capabilities and tools to manage and analyze large quantities and different types of data from multiple sources. Easy-to-use query and reporting tools for casual business users and more sophisticated analytical toolsets for power users are included.

that's been restructured and reorganized for reporting and analysis. Hadoop clusters pre-process big data for use in the data warehouse, data marts, or an analytic platform, or for direct querying by power users. Outputs include reports and dashboards as well as query results. Chapter 12 discusses the various types of BI users and BI reporting in greater detail.

ANALYTICAL TOOLS: RELATIONSHIPS, PATTERNS, TRENDS

Once data have been captured and organized using the business intelligence technologies we have just described, they are available for further analysis using software for database querying and reporting, multidimensional data analysis (OLAP), and data mining. This section will introduce you to these tools, with more detail about business intelligence analytics and applications in Chapter 12.

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.

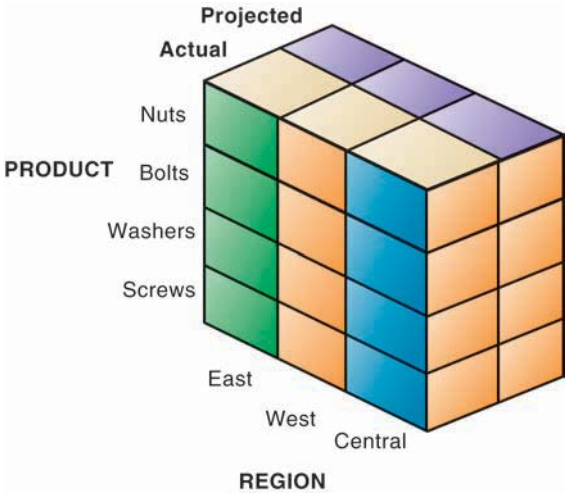
Figure 6.13 shows a multidimensional model that could be created to represent products, regions, actual sales, and projected sales. A matrix of actual sales can be stacked on top of a matrix of projected sales to form a cube with six faces. If you rotate the cube 90 degrees one way, the face showing 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 2013?" 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

FIGURE 6.13 MULTIDIMENSIONAL DATA MODEL



This view shows 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.

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. 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.

Caesars Entertainment, formerly known as Harrah's Entertainment, is the largest gaming company in the world. It continually analyzes data about its customers gathered when people play its slot machines or use its casinos and hotels. The corporate marketing department uses this information to build a detailed gambling profile, based on a particular customer's ongoing value to the company. For instance, data mining lets Caesars know the favorite gaming experience of a regular customer at one of its 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 improved Caesars's profits so much that it became the centerpiece of the firm's business strategy.

Text Mining and Web Mining

However, unstructured data, most in the form of text files, is believed to account for over 80 percent of useful organizational information and is one

of the major sources of big data that firms want to analyze. E-mail, memos, call center transcripts, survey responses, legal cases, patent descriptions, and service reports are all valuable for finding patterns and trends that will help employees make better business decisions. **Text mining** tools are now available to help businesses analyze these data. These tools are able to extract key elements from unstructured big data sets, discover patterns and relationships, and summarize the information.

Businesses might turn to text mining to analyze transcripts of calls to customer service centers to identify major service and repair issues or to measure customer sentiment about their company. **Sentiment analysis** software is able to mine text comments in an e-mail message, blog, social media conversation, or survey form to detect favorable and unfavorable opinions about specific subjects.

For example, the discount broker Charles Schwab uses Attensity Analyze software to analyze hundreds of thousands of its customer interactions each month. The software analyzes Schwab's customer service notes, e-mails, survey responses, and online discussions to discover signs of dissatisfaction that might cause a customer to stop using the company's services. Attensity is able to automatically identify the various "voices" customers use to express their feedback (such as a positive, negative, or conditional voice) to pinpoint a person's intent to buy, intent to leave, or reaction to a specific product or marketing message. Schwab uses this information to take corrective actions such as stepping up direct broker communication with the customer and trying to quickly resolve the problems that are making the customer unhappy.

The Web is another rich source of unstructured big data for revealing patterns, trends, and insights into customer behavior. The discovery and analysis of useful patterns and information from the World Wide Web is called **Web mining**. Businesses might turn to Web mining to help them understand customer behavior, evaluate the effectiveness of a particular Web site, or quantify the success of a marketing campaign. For instance, marketers use the Google Trends and Google Insights for Search services, which track the popularity of various words and phrases used in Google search queries, to learn what people are interested in and what they are interested in buying.

Web mining looks for patterns in data through content mining, structure mining, and usage mining. Web content mining is the process of extracting knowledge from the content of Web pages, which may include text, image, audio, and video data. Web structure mining examines data related to the structure of a particular Web site. For example, links pointing to a document indicate the popularity of the document, while links coming out of a document indicate the richness or perhaps the variety of topics covered in the document. Web usage mining examines user interaction data recorded by a Web server whenever requests for a Web site's resources are received. The usage data records the user's behavior when the user browses or makes transactions on the Web site and collects the data in a server log. Analyzing such data can help companies determine the value of particular customers, cross marketing strategies across products, and the effectiveness of promotional campaigns.

The Interactive Session on Technology describes organizations' experiences as they use the analytical tools and business intelligence technologies we have described to grapple with "big data" challenges.

INTERACTIVE SESSION: TECHNOLOGY

BIG DATA, BIG REWARDS

Today's companies are dealing with an avalanche of data from social media, search, and sensors as well as from traditional sources. In 2012, the amount of digital information generated is expected to reach 988 exabytes, which is the equivalent to a stack of books from the sun to the planet Pluto and back. Making sense of "big data" has become one of the primary challenges for corporations of all shapes and sizes, but it also represents new opportunities. How are companies currently taking advantage of big data opportunities?

The British Library had to adapt to handle big data. Every year visitors to the British Library Web site perform over 6 billion searches, and the library is also responsible for preserving British Web sites that no longer exist but need to be preserved for historical purposes, such as the Web sites for past politicians. Traditional data management methods proved inadequate to archive millions of these Web pages, and legacy analytics tools couldn't extract useful knowledge from such quantities of data. So the British Library partnered with IBM to implement a big data solution to these challenges. IBM BigSheets is an insight engine that helps extract, annotate, and visually analyze vast amounts of unstructured Web data, delivering the results via a Web browser. For example, users can see search results in a pie chart. IBM BigSheets is built atop the Hadoop framework, so it can process large amounts of data quickly and efficiently.

State and federal law enforcement agencies are analyzing big data to discover hidden patterns in criminal activity such as correlations between time, opportunity, and organizations, or non-obvious relationships (see Chapter 4) between individuals and criminal organizations that would be difficult to uncover in smaller data sets. Criminals and criminal organizations are increasingly using the Internet to coordinate and perpetrate their crimes. New tools allow agencies to analyze data from a wide array of sources and apply analytics to predict future crime patterns. This means that law enforcement can become more proactive in its efforts to fight crime and stop it before it occurs.

In New York City, the Real Time Crime Center data warehouse contains millions of data points on city crime and criminals. IBM and the New York City Police Department (NYPD) worked together to create the warehouse, which contains data on over 120 million criminal complaints, 31 million national crime records, and 33 billion public records. The system's search capabilities allow the NYPD to quickly obtain data from any of these data sources. Information on criminals, such as a suspect's photo with details of past offenses or addresses with maps, can be visualized in seconds on a video wall or instantly relayed to officers at a crime scene.

Other organizations are using the data to go green, or, in the case of Vestas, to go even greener. Headquartered in Denmark,Vestas is the world's largest wind energy company, with over 43,000 wind turbines across 66 countries. Location data are important to Vestas so that it can accurately place its turbines for optimal wind power generation. Areas without enough wind will not generate the necessary power, but areas with too much wind may damage the turbines. Vestas relies on location-based data to determine the best spots to install their turbines.

To gather data on prospective turbine locations, Vestas's wind library combines data from global weather systems along with data from existing turbines. The company's previous wind library provided information in a grid pattern, with each grid measuring 27 x 27 kilometers (17 x 17 miles). Vestas engineers were able to bring the resolution down to about 10 x 10 meters (32 x 32 feet) to establish the exact wind flow pattern at a particular location. To further increase the accuracy of its turbine placement models, Vestas needed to shrink the grid area even more, and this required 10 times as much data as the previous system and a more powerful data management platform.

The company implemented a solution consisting of IBM InfoSphere BigInsights software running on a high-performance IBM System x iDataPlex server. (InfoSphere BigInsights is a set of software tools for big data analysis and visualization, and is powered by Apache Hadoop.) Using these technologies, Vestas in creased the size of its wind library and is able manage and analyze location and weather data with models that are much more powerful and precise.

Vestas's wind library currently stores 2.8 petabytes of data and includes approximately 178 parameters, such as barometric pressure, humidity, wind direction, temperature, wind velocity, and other company historical data. Vestas plans to add global deforestation metrics, satellite images, geospatial data, and data on phases of the moon and tides.

The company can now reduce the resolution of its wind data grids by nearly 90 percent, down to a 3 x 3 kilometer area (about 1.8 x 1.8 miles). This capability enables Vestas to forecast optimal turbine placement in 15 minutes instead of three weeks, saving a month of development time for a turbine site and enabling Vestas customers to achieve a return on investment much more quickly.

Companies are also using big data solutions to analyze consumer sentiment. For example, car-rental giant Hertz gathers data from Web surveys, e-mails, text messages, Web site traffic patterns, and data generated at all of Hertz's 8,300 locations in 146 countries. The company now stores all of that data centrally instead of within each branch, reducing time spent processing data and improving company response time to customer feedback and changes in sentiment. For example, by analyzing data generated from multiple sources, Hertz was able to determine that delays were occurring for returns in Philadelphia during specific times of the day. After investigating this anomaly, the company was able to quickly adjust staffing levels at its Philadelphia office during those peak times, ensuring a manager was present to resolve any issues. This enhanced Hertz's performance and increased customer satisfaction.

There are limits to using big data. Swimming in numbers doesn't necessarily mean that the right information is being collected or that people will make smarter decisions. Last year, a McKinsey Global Institute report cautioned there is a shortage of specialists who can make sense of all the information being generated. Nevertheless, the trend towards big data shows no sign of slowing down; in fact, it's much more likely that big data is only going to get bigger.

Sources: Samuel Greengard," Big Data Unlocks Business Value," Baseline, January 2012; Paul S. Barth, "Managing Big Data: What Every CIO Needs to Know," *CIO Insight*, January 12, 2012; IBM Corporation, "Vestas: Turning Climate into Capital with Big Data," 2011; IBM Corporation, "Extending and enhancing law enforcement capabilities," "How Big Data Is Giving Hertz a Big Advantage," and "British Library and J Start Team Up to Archive the Web," 2010.

CASE STUDY QUESTIONS

1. Describe the kinds of big data collected by the organizations described in this case.
2. List and describe the business intelligence technologies described in this case.
3. Why did the companies described in this case need to maintain and analyze big data? What business benefits did they obtain?
4. Identify three decisions that were improved by using big data.
5. What kinds of organizations are most likely to need big data management and analytical tools? Why?

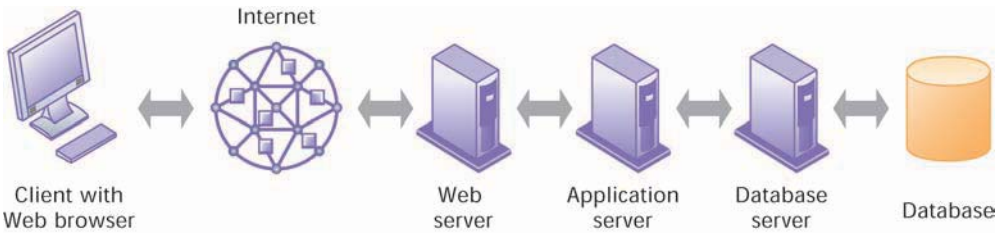
DATABASES AND THE WEB

Have you ever tried to use the Web to place an order or view a product catalog? If so, you were probably using a Web site linked to an internal corporate database. Many companies now use the Web to make some of the information in their internal databases available to customers and business partners.

Suppose, for example, a customer with a Web browser wants to search an online retailer's database for pricing information. Figure 6.14 illustrates how that customer might access the retailer's internal database over the Web. The user accesses the retailer's Web site over the Internet using Web browser software on his or her client PC. The user's Web browser software requests 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 passes these requests for data to software that translates HTML commands into SQL so the commands can be processed by the DBMS working with the database. In a client/server environment, the

FIGURE 6.14 LINKING INTERNAL DATABASES TO THE WEB



Users access an organization's internal database through the Web using their desktop PCs and Web browser software.

DBMS resides on a dedicated computer called a **database server**. The DBMS receives the SQL requests and provides the required data. Middleware transfers information from the organization's internal database back to the Web server for delivery in the form of a Web page to the user.

Figure 6.14 shows that the middleware working between the Web server and the DBMS is an application server running on its own dedicated computer (see Chapter 5). 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, Web browser software is much easier to use than proprietary query tools. Second, the Web interface requires few or no changes to the internal database. 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.

Accessing corporate databases through the Web is creating new efficiencies, opportunities, and business models. ThomasNet.com provides an up-to-date online directory of more than 650,000 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.

Other companies have created entirely new businesses based on access to large databases through the Web. One is the social networking service Facebook, which helps users stay connected with each other and meet new people. Facebook features "profiles" with information on more than 950 million active users with information about themselves, including interests, friends, photos, and groups with which they are affiliated. Facebook maintains a massive database to house and manage all of this content.

There are also many Web-enabled databases in the public sector to help consumers and citizens access helpful information. The Interactive Session on Organizations describes one of these databases, which has generated controversy over its methods for providing consumer product safety data.

INTERACTIVE SESSION: ORGANIZATIONS

CONTROVERSY WHIRLS AROUND THE CONSUMER PRODUCT SAFETY DATABASE

Michele Witte was one of dozens of parents who lost their children because of the defective design of drop-side cribs. In 1997, Witte's 10-month-old son Tyler perished when the drop-side rail on his crib came loose, partially detached, and then trapped his neck between the rail and the headboard. The cribs are now banned. Witte wishes that a public information resource for consumer complaints had been available prior to the death of her child. Reading other parents' horror stories might have dissuaded her from purchasing a drop-side crib.

In March 2011, the U.S. Consumer Product Safety Commission (CPSC) stood poised to meet the needs of parents like Witte by launching an online database, located at www.saferproducts.gov. The database will provide the public with access to the full repository of product safety complaints that it has received. Users can submit these complaints online directly into the database. Visitors to the database will be able to search for products, read other complaints, and view safety warnings issued by the CPSC. Complaints in the database will include a description of product, the harm or risk from the product, the name of the manufacturer, contact information, and an affirmation that the submitter is telling the truth. The submitter's name will not appear in the database but could be provided to manufacturers if the submitter agreed.

Consumer advocates such as the Consumer Federation of America are praising the database as a revolutionary resource that will drastically improve the way consumers buy products. However, manufacturing companies and many members of Congress are in opposition. They argue that because any user can submit a complaint, the database will be filled with inaccurate and misleading information—"fictitious slams" against products. It will also be open to abuse from customers with an axe to grind, or trial lawyers seeking to tarnish a product or manufacturer's reputation for personal gain.

The database represents an increase in visibility and authority for the CPSC, which was formed in 1972 by the Consumer Product Safety Act. The role of the CPSC is to regulate thousands of different types of products, with special focus on those that are not regulated by other areas of the government already, like food, firearms, and automobiles. (The CPSC database does not include safety problems with these products.)

The CPSC collects reports on defective products from consumers, health care providers, death certificates, media accounts, and other sources. It uses that information to make decisions on product recalls and bans, but until recently, very little of that information was accessible to the public. Federal law formerly required the approval of manufacturers to publicize that information, and manufacturers weren't eager to release information about their faulty products. Not only that, but the CPSC had to negotiate directly with manufacturers to determine the terms of product recalls. Because this process usually takes a year or more, consumers continue to buy shoddy and perhaps dangerous products like drop-side cribs in the interim.

Under the new system, complaints filed by consumers will be posted online and be available to the public within 15 days. Companies will be notified within 5 days when complaints are made about their products, and the CPSC will give them 10 days to respond publicly and have their comments published alongside the complaints in the database. Users will have the option for their comments to remain confidential if they prefer. Manufacturers will be able to appeal to the CPSC to eliminate false or misleading complaints, and complaints will be limited to defects that can cause injury, not reliability or product quality.

At a time when the federal budget is under increased scrutiny, programs like the CPSC database have become targets for cost-cutting, and manufacturers have seized an opportunity to stop the database in its tracks. The law gave CPSC new authority to regulate unsafe products but businesses say it is overly burdensome. A House Energy and Commerce subcommittee is considering draft legislation to restrict who can submit reports to the database, to improve how products are identified, and to resolve claims that reports are inaccurate.

Despite strong opposition from manufacturers and others, in March 2011, the site was launched to generally positive reviews. The CPSC provided additional features, like the ability to attach images to comments. Commenters must provide their name, mailing address, telephone number, and e-mail address, which is expected to curtail the types of anonymous

comments that manufacturers fear. Even so, keeping the database free of inaccurate reports is likely to require more time and hours than the CPSC staff will be able to provide.

Since the database went live, there have been hundreds of thousands of visits to the site and millions of product searches conducted by visitors, according to the Consumer Product Safety commission. Despite its growing popularity, it may not survive congressional attempts to take away its funding, in response to pressures to reduce the federal budget as well as criticism from the business community. Time will tell whether saferproducts.gov becomes an indispensable consumer resource.

Sources: www.SaferProducts.gov, accessed May 22, 2012; Josh Cable, "Democrats Defend Consumer Product Safety Database," *Industry Week,* July 7, 2011; Don Mays, "My Experience With the CPSC Database," blogs.consumerreports.com, March 16, 2011; Andrew Martin, "Child-Product Makers Seek to Soften New Rules," *The New York Times*, February 21, 2011; Lyndsey Layton, "Consumer Product Safety Commission to Launch Public Database of Complaints," *Washington Post*, January 10, 2011; Jayne O'Donnell, "Product-Safety Database Under Multiple Attacks," *USA Today*, April 12, 2011.

CASE STUDY QUESTIONS

1. What is the value of the CPSC database to consumers, businesses, and the U.S. government?
2. What problems are raised by this database? Why is it so controversial? Why is data quality an issue?
3. Name two entities in the CPSC database and describe some of their attributes.
4. When buying a crib, or other consumer product for your family, would you use this database? Why or why not?

6.4 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 policy lays 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.

You may hear the term **data governance** used to describe many of these activities. Promoted by IBM, data governance deals with the policies and processes for managing the availability, usability, integrity, and security of the data employed in an enterprise, with special emphasis on promoting privacy, security, data quality, and compliance with government regulations.

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 or your sales system and inventory system showed different prices for the same product? Data that are inaccurate, untimely, or inconsistent with other sources of information lead to incorrect decisions, product recalls, and financial losses. Gartner Inc. 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. A Sirius Decisions study on "The Impact of Bad Data on Demand Creation" found that 10 to 25 percent of customer and prospect records contain critical data errors. Correcting these errors at their source and following best practices for promoting data quality increased the productivity of the sales process and generated a 66 percent increase in revenue.

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.

Think of all the times you've 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.

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.

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 errors 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.

Data quality problems are not just business problems. They also pose serious problems for individuals, affecting their financial condition and even their jobs. For example, inaccurate or outdated data about consumers' credit histories maintained by credit bureaus can prevent creditworthy individuals from obtaining loans or lower their chances of finding or keeping a job.

Learning Track Modules

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
3. Hierarchical and Network Data Models

Review Summary

1. *What are the problems of managing data resources in a traditional file environment and how are they solved by a database management system?*

   Traditional file management techniques make it difficult for organizations to keep track of all of the pieces of data they use in a systematic way and to organize these data so that they can be easily accessed. Different functional areas and groups were allowed to develop their own files independently. Over time, this traditional file management environment creates problems such as data redundancy and inconsistency, program-data dependence, inflexibility, poor security, and lack of data sharing and availability. A database management system (DBMS) solves these problems with software that permits centralization of data and data management so that businesses have a single consistent source for all their data needs. Using a DBMS minimizes redundant and inconsistent files.

2. *What are the major capabilities of DBMS and why is a relational DBMS so powerful?*

   The principal capabilities of a DBMS includes a data definition capability, a data dictionary capability, and a data manipulation language. The data definition capability specifies the structure and content of the database. The data dictionary is an automated or manual file that stores information about the data in the database, including names, definitions, formats, and descriptions of data elements. The data manipulation language, such as SQL, is a specialized language for accessing and manipulating the data in the database.

   The relational database has been the primary method for organizing and maintaining data in information systems because it is so flexible and accessible. It organizes data in two-dimensional tables called relations with rows and columns. 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. Relational database tables can be combined easily to deliver data required by users, provided that any two tables share a common data element. Non-relational databases are becoming popular for managing types of data that can't be handled easily by the relational data model. Both relational and non-relational database products are available as cloud computing services.

3. *What are some important database design principles?*

   Designing a database requires both a logical design and a physical design. The logical design models the database from a business perspective. The organization's data model should reflect its key business processes and decision-making requirements. The process of creating small, stable, flexible, and adaptive data structures from complex groups of data when designing a relational database is termed normalization. A well-designed relational database will not have many-to-many relationships, and all attributes for a specific entity will only apply to that entity. It will try to enforce referential integrity rules to ensure that relationships between coupled tables remain consistent. An entity-relationship diagram graphically depicts the relationship between entities (tables) in a relational database.

4. *What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?*

   Contemporary data management technology has an array of tools for obtaining useful information from all the different types of data used by businesses today, including semi-structured and unstructured big data in vast quantities. These capabilities include data warehouses and data marts, Hadoop, in-memory computing, and analytical platforms. OLAP represents relationships among data as a multidimensional structure, which can be visualized as cubes of data and cubes within cubes of data, enabling more sophisticated data analysis. 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. Text mining tools help businesses analyze large unstructured data sets consisting of text. Web mining tools focus on analysis of useful patterns and information from the World Wide Web, examining the structure of Web sites and activities of Web site users as well as the contents of Web pages. Conventional databases can be linked via middleware to the Web or a Web interface to facilitate user access to an organization's internal data.

5. *Why are information policy, data administration, and data quality assurance essential for managing the firm's 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.

   Data that are inaccurate, incomplete, or inconsistent create serious operational and financial problems for businesses because they may create inaccuracies in product pricing, customer accounts, and inventory data, and lead to inaccurate decisions about the actions that should be taken by the firm. Firms must take special steps to make sure they have a high level of data quality. These include using enterprise-wide data standards, databases designed to minimize inconsistent and redundant data, data quality audits, and data cleansing software.

Key Terms

Analytic platform, 256
Attribute, 241
Big data, 254
Bit, 241
Byte, 241
Data administration, 265
Data cleansing, 267
Data definition, 249
Data dictionary, 249
Data governance, 266
Data inconsistency, 243
Data manipulation language, 250
Data mart, 255
Data mining, 258
Data quality audit, 267
Data redundancy, 243
Data warehouse, 255
Database, 244
Database administration, 266
Database management system (DBMS), 244
Database server, 263
Entity, 241
Entity-relationship diagram, 253
Field, 241
File, 241
Foreign key, 247
Hadoop, 255
In-memory computing, 256
Information policy, 265
Key field, 247
Non-relational database management systems, 247
Normalization, 252
Online analytical processing (OLAP), 257
Primary key, 247
Program-data dependence, 244
Record, 241
Referential integrity, 253
Relational DBMS, 246
Sentiment analysis, 260
Structured Query Language (SQL), 250
Text mining, 260
Tuple, 247
Web mining, 260

Review Questions

1. What are the problems of managing data resources in a traditional file environment and how are they solved by a database management system?
   - List and describe each of the components in the data hierarchy.
   - Define and explain the significance of entities, attributes, and key fields.
   - List and describe the problems of the traditional file environment.
   - Define a database and a database management system and describe how it solves the problems of a traditional file environment.
2. What are the major capabilities of DBMS and why is a relational DBMS so powerful?
   - Name and briefly describe the capabilities of a DBMS.
   - Define a relational DBMS and explain how it organizes data.

- List and describe the three operations of a relational DBMS.
- Explain why non-relational databases are useful.

3. What are some important database design principles?

- Define and describe normalization and referential integrity and explain how they contribute to a well-designed relational database.
- Define and describe an entity-relationship diagram and explain its role in database design.

4. What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

- Define big data and describe the technologies for managing and analyzing it.
- List and describe the components of a contemporary business intelligence infrastructure.
- Describe the capabilities of online analytical processing (OLAP).
- Define data mining, describing how it differs from OLAP and the types of information it provides.
- Explain how text mining and Web mining differ from conventional data mining.
- Describe how users can access information from a company's internal databases through the Web.

5. Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?

- Describe the roles of information policy and data administration in information management.
- Explain why data quality audits and data cleansing are essential.

Discussion Questions

1. It has been said there is no bad data, just bad management. Discuss the implications of this statement.

2. To what extent should end users be involved in the selection of a database management system and database design?

3. What are the consequences of an organization not having an information policy?

Hands-On MIS Projects

The projects in this section give you hands-on experience in analyzing data quality problems, establishing company-wide data standards, creating a database for inventory management, and using the Web to search online databases for overseas business resources.

Management Decision Problems

1. Emerson Process Management, a global supplier of measurement, analytical, and monitoring instruments and services based in Austin, Texas, had a new data warehouse designed for analyzing customer activity to improve service and marketing. However, the data warehouse was full of inaccurate and redundant data. The data in the warehouse came from numerous transaction processing systems in Europe, Asia, and other locations around the world. The team that designed the warehouse had assumed that sales groups in all these areas would enter customer names and addresses the same way. In fact, companies in different countries were using multiple ways of entering quote, billing, shipping,

and other data. Assess the potential business impact of these data quality problems. What decisions have to be made and steps taken to reach a solution?

2. Your industrial supply company wants to create a data warehouse where management can obtain a single corporate-wide view of critical sales information to identify bestselling products, key customers, and sales trends. Your sales and product information are stored in several different systems: a divisional sales system running on a Unix server and a corporate sales system running on an IBM mainframe. You would like to create a single standard format that consolidates these data from both systems. In MyMISLab, you can review the proposed format, along with sample files from the two systems that would supply the data for the data warehouse. Then answer the following questions:

   - What business problems are created by not having these data in a single standard format?
   - How easy would it be to create a database with a single standard format that could store the data from both systems? Identify the problems that would have to be addressed.
   - Should the problems be solved by database specialists or general business managers? Explain.
   - Who should have the authority to finalize a single company-wide format for this information in the data warehouse?

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. Using the information found in the tables in MyMISLab, build a simple relational database to manage information about Sylvester's suppliers and products. 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. For each supplier, 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

This project develops skills in searching Web-enabled databases with information about products and services in faraway locations.

Your company is located in Greensboro, North Carolina, and manufactures office furniture of various types. You are considering opening a facility to manufacture and sell your products in Australia. You would like to contact organizations that offer many services necessary for you to open your Australian office and manufacturing facility, including lawyers, accountants, import-export experts, and telecommunications equipment and support firm. Access the following online databases to locate companies that you would like to meet with during your upcoming trip: Australian Business Register (abr.gov.au), AustraliaTrade Now

(australiatradenow.com), and the Nationwide Business Directory of Australia (www.nationwide.com.au). If necessary, use search engines such as Yahoo and Google.

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

Video Cases

Video Cases and Instructional Videos illustrating some of the concepts in this chapter are available. Contact your instructor to access these videos.

Collaboration and Teamwork Project

In MyMISLab, you will find a Collaboration and Teamwork Project dealing with the concepts in this chapter. You will be able to use Google Sites, Google Docs, and other open source collaboration tools to complete the assignment.

Lego: Embracing Change by Combining BI with a Flexible Information System

CASE STUDY

The Lego Group, which is headquartered in Billund, Denmark, is one of the largest toy manufacturers in the world. Lego's main products have been the bricks and figures that children have played with for generations. The Danish company has experienced sustained growth since its founding in 1932, and for most of its history its major manufacturing facilities were located in Denmark.

In 2003, Lego was facing tough competition from imitators and manufacturers of electronic toys. In an effort to reduce costs, the group decided to initiate a gradual restructuring process that continues today. In 2006, the company announced that a large part of its production would be outsourced to the electronics manufacturing service company Flextronics, which has plants in Mexico, Hungary, and the Czech Republic. The decision to outsource production came as a direct consequence of an analysis of Lego's total supply chain. To reduce labor costs, manually intensive processes were outsourced, keeping only the highly skilled workers in Billund. Lego's workforce was gradually reduced from 8,300 employees in 2003 to approximately 4,200 in 2010. Additionally, production had to be relocated to places closer to its natural markets. As a consequence of all these changes, Lego transformed itself from a manufacturing firm to a market-oriented company that is capable of reacting fast to changing global demand.

Lego's restructuring process, coupled with double-digit sales growth in the past few years, has led to the company's expansion abroad and made its workforce more international. These changes presented supply chain and human resources challenges to the company. The supply chain had to be reengineered to simplify production without reducing quality. Improved logistics planning allowed Lego to work more closely with retailers, suppliers, and the new outsourcing companies. At the same time, the human resources (HR) department needed to play a more strategic role inside the company. HR was now responsible for implementing effective policies aimed at retaining and recruiting the most qualified employees from a diversity of cultural backgrounds.

Adapting company operations to these changes required a flexible and robust IT infrastructure with business intelligence capabilities that could help management perform better forecasting and planning. As part of the solution, Lego chose to move to SAP business suite software. SAP AG, a German company that specializes in enterprise software solutions, is one of the leading software companies in the world. SAP's software products include a variety of applications designed to efficiently support all of a company's essential functions and operations. Lego chose to implement SAP's Supply Chain Management (SCM), Product Lifecycle Management (PLM), and Enterprise Resources Planning (ERP) modules.

The SCM module includes essential features such as supply chain monitoring and analysis as well as forecasting, planning, and inventory optimization. The PLM module enables managers to optimize development processes and systems. The ERP module includes, among other applications, the Human Capital Management (HCM) application for personnel administration and development.

SAP's business suite is based on a flexible three-tier client-server architecture that can easily be adapted to the new Service-Oriented Architecture (SOA) available in the latest versions of the software. In the first tier, a client interface—a browser-type graphical user interface (GUI) running on either a laptop, desktop, or mobile device—submits users' requests to the application servers. The applications servers—the second tier in the system—receive and process clients' requests. In turn, these application servers send the processed requests to the database system—the third tier—which consists of one or more relational databases. SAP's business suite supports databases from different vendors, including those offered by Oracle, Microsoft, MySQL, and others. The relational databases contain the tables that store data on Lego's products, daily operations, the supply chain, and thousands of employees. Managers can easily use the SAP query tool to obtain reports from the databases, because it does not require any technical skill. Additionally, the distributed architecture enables authorized personnel to have direct access to the database system from the company's various locations, including those in Europe, North America, and Asia.

SAP's ERP-HCM module includes advanced features such as "Talent Manager" as well those for handling employee administration, reporting, and travel and time management. These features allow Lego's HR personnel to select the best candidates, schedule their training, and create a stimulus plan to retain

them. It is also possible to include performance measurements and get real-time insight into HR trends. Using these advanced features, together with tools from other software vendors, Lego's managers are able to track employees' leadership potential, develop their careers, and forecast the recruiting of new employees with certain skills. n.

Sources: "Business 2010: Embracing the Challenge of Change," The Economist Intelligence Unit, February 2005 (http://graphics.eiu.com/files/ad_pdfs/Business%202010_Global_ FINAL.pdf, accessed November 16, 2010); "Lego Creates Model Business Success with SAP and IBM," IBM Global Financing, May 19, 2010 (www-01.ibm.com/software/success/cssdb.nsf/CS/STRD-85KGS6?OpenDocument, October 20, 2010); "Human Resources as an Exponent of Good Governance" (in Danish) (www.sat.com, October 20, 2010); "Lego, The Toy of the Century Had to Reinvent the Supply-Chain to Save the Company," Supply Chain Digest, September 25, 2007 (www.scdigest.com/assets/on_target/07-09-25-7.php?cid=1237, accessed November 16, 2010); G. W. Anderson, T. Rhodes, J. Davis, and J. Dobbins, SAMS Teach Yourself SAP in 24 hours (Indianapolis, IN: SAMS, 2008).

CASE STUDY QUESTIONS

1. Explain the role of the database in SAP's three-tier system.
2. Explain why distributed architectures are flexible.
3. Identify some of the business intelligence features included in SAP's business software suite.
4. What are the main advantages and disadvantages of having multiple databases in a distributed architecture? Explain.

Case contributed by Daniel Ortiz Arroyo, Aalborg University