

# CHAPTER 6

---

## 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?

### OPENING CASE: CAN HP MINE SUCCESS FROM AN ENTERPRISE DATA WAREHOUSE?

The essential message of this chapter is the statement that “Organizations need to manage their data assets very carefully to make sure that the data are easily accessed and used by managers and employees across the organization.” Data have now become central and even vital to an organization's survival. You can illustrate these comments by referencing the opening case, “Can HP Mine Success from an Enterprise Data Warehouse?,” in order to stress the importance of data and database systems for success in business.

Without a consistent view of the enterprise, HP senior executives struggled with decisions on matters such as the size of sales and service teams assigned to particular systems. HP had too many different information system applications in too many computer centers. It had too many different database technologies and way too many

different databases. As with most organizations, departments were allowed to create, manage and use their own databases without regard towards sharing the data with other departments—*islands of information* at their finest. HP CIO Randy Mott began consolidating hundreds of data marts into a single data warehouse. He had three goals for the database: it had to always be up-to-date, consistent for the entire enterprise, and complete. The Neoview system includes all of the data used by a company and not just partial segments of data or the company.

What's interesting and intriguing about the opening vignette is how it points out that every organization, even a technology company like HP, struggles with the need to manage data and information as an important resource. How businesses store, organize, and manage their data has a tremendous impact on organizational effectiveness. Companies need to manage their data to help them reduce costs, improve operational efficiency and decision making, and most of all, boost profitability.

**6.1****ORGANIZING DATA IN A TRADITIONAL  
FILE ENVIRONMENT**

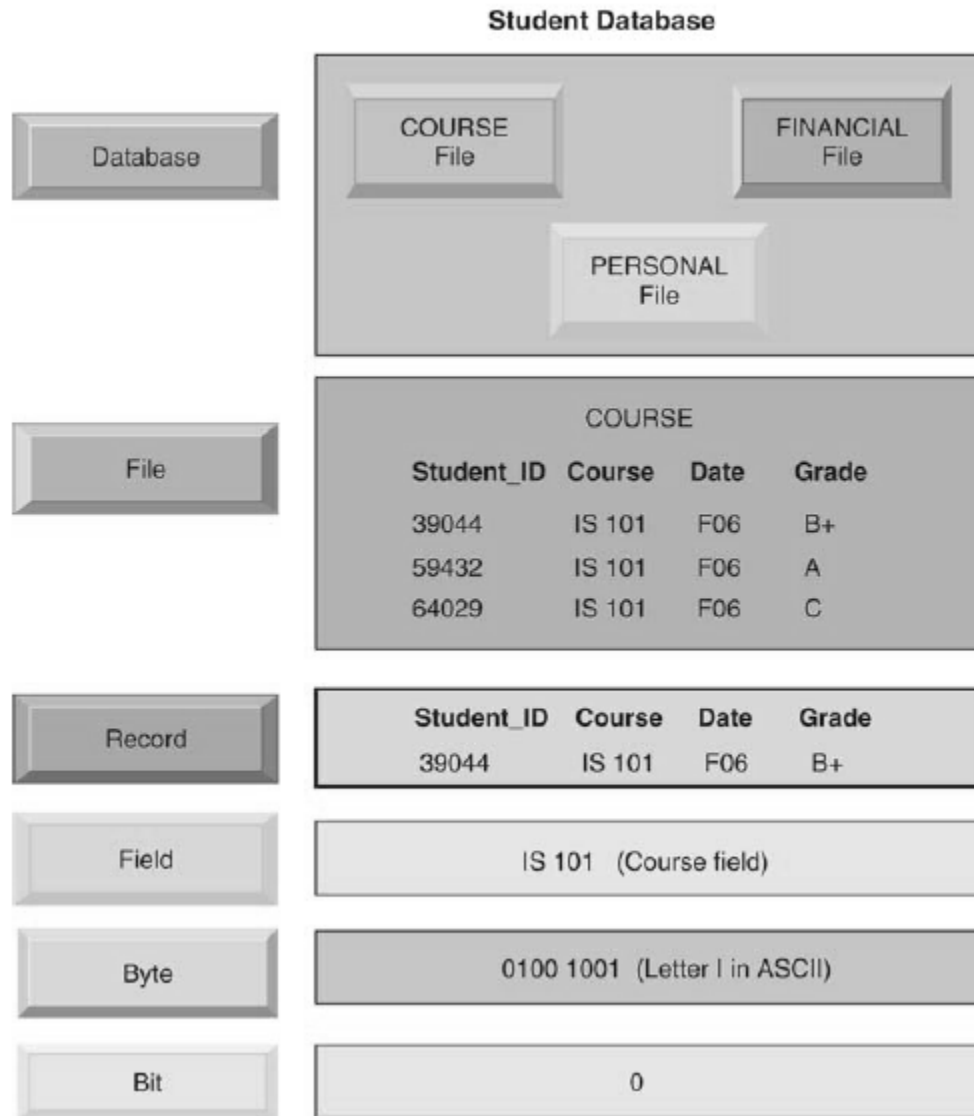
---

Information is becoming as important a business resource as money, material, and people. Even though a company compiles millions of pieces of data doesn't mean it can produce information that its employees, suppliers, and customers can use. Businesses are realizing the competitive advantage they can gain by compiling useful information, not just data.

Why should students learn about organizing data? Because it's almost inevitable that someday they will be establishing or at least working with a database of some kind. Understanding the jargon is the first step to understanding the whole concept of managing and maintaining information.

**FILE ORGANIZATION TERMS AND CONCEPTS**

The first few terms, **field**, **record**, **file**, **database**, are depicted in Figure 6-1, which shows the relationship between them.

**FIGURE 6-1 THE DATA HIERARCHY**

An **entity** is basically the person, place, thing, or event on which we maintain information. Each characteristic or quality describing an entity is called an **attribute**.

## PROBLEMS WITH THE TRADITIONAL FILE ENVIRONMENT

Many problems such as data redundancy, data inconsistency, program-data dependence, inflexibility, poor data security, and of data sharing and availability among applications have occurred with traditional file environments.

### Data Redundancy and Inconsistency

Students will recognize the situation that occurs when they move and change addresses. You can use the example of the university. A student notifies everyone of their new address including the registrar's office. Everything is going smoothly until, at the end of the year, the library sends an overdue notice to the old address. Why? Because the new address was changed in one database, but the library maintains a separate database and the address was never changed in it. This is an example of **data inconsistency**. It also illustrates **data redundancy**. That is, the information is in two separate databases with duplicate records. This wastes storage resources, but more important it also leads to incorrect data as in the example above.

### **Program-Data Dependence**

Even more troublesome is when several departments or individuals decide to set up their own islands of information. This usually happens because they find the main system inflexible or it just doesn't fit their needs. So they set up their own fields and records and files and use them in their own programs to manipulate data according to their needs. Now each department is spending dollars and time to establish and maintain islands of information because of **program-data dependence**.

Taking this problem even further, the fields and records for marketing probably don't have the same structure and meaning as the fields and records for accounting, or those for production. Each record describes basically the same entity (customers or products), but it is very possible that each database file will have different information, or attributes, in records concerning the same entity.

All of this may happen with the best of intentions. All departments began with the goal of making their part of the organization more efficient. Eventually these good intentions can cost big dollars to bring the islands together, resolve data conflicts, and retrain people to understand the new database structures.

### **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.

### **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**

Pieces of information in different files and different parts of the organization cannot be related to one another. This situation makes it virtually impossible for information to be shared or accessed in a timely manner. For example, assume that the marketing department has a promotional program whereby all individuals placing an order last month of \$1 000 or more are issued a \$50 gift card on their next purchase. After placing their order, the customer later returns the product and the accounting department issues

them a full credit. If the marketing department maintains their own customer sales database, then they would have no way of knowing that the customer returned the product. Thus, even customers who returned their orders would still be issued the \$50 gift card. In this example you can see the necessity for data sharing and making it available to the different parts of the organization that need to access it.

**Bottom Line: Managers and workers must know and understand how databases are constructed so they know how to use the information resource to their advantage. Managers must guard against problems inherent with islands of information and understand that sometimes resolution of short-term problems is far costlier in the long term.**

## 6.2

## THE DATABASE APPROACH TO DATA MANAGEMENT

The key to establishing an effective, efficient **database** is to involve the entire organization as much as possible, even if everyone will not immediately be connected to it or use it. Perhaps they won't be a part of it in the beginning, but they very well could be later on.

### DATABASE MANAGEMENT SYSTEMS

A **database management system (DMBS)** 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.

*Physical views* of items are often different from the *logical views* of the same items when they are actually being used. For instance, assume you store tablets of paper in your lower-right desk drawer. You store your pencils in the upper-left drawer. When it comes time to write a note, you pull out the paper and pencil and put them together on your desktop. It isn't important to the task at hand where the items were stored *physically*; you are concerned with the *logical* idea of the two items coming together to help you accomplish the task.

The physical view of data focuses on where the data are actually stored in the record or in a file. The physical view is important to programmers who must manipulate the data as they are *physically stored* in the database.

Does it really matter to the user that the customer address is physically stored on the disk before the customer name? Probably not. However, when users create a report of customers located in Nova Scotia, they generally will list the customer name first and

then the address. So it's more important to the end user to bring the data from its physical location on the storage device to a *logical* view in the output device, whether screen or paper.

### 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. 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 enables the organization to centrally manage data, their use, and security.

### Relational DBMS

The most popular type of DBMS today for PCs as well as for larger computers and mainframes is the relational DBMS. A **relational DBMS** uses tables in which data are stored to extract and combine data in whatever form or format the user needs. The tables are sometimes called files, although that is actually a misnomer, since you can have multiple tables in one file.

Ask students to think about designing a database, for an example like a newspaper delivery business. In order to succeed, you need to keep accurate, useful information for each customer. In the database, the data about a single customer resides in a row in a customer table. Rows are commonly referred to as **records**, or in a very technical term, a **tuple**. Thus, for each customer, you create a record. Within each record you have the following fields: customer name (ask students to consider why the first name and second name should be separate fields), address (again, what elements should be made separate fields?), ID, date last paid. Smith, Jones, and Brooks are the records within a file you decide to call Paper Delivery. The entities then are Smith, Jones, and Brooks, the people about whom you are maintaining information. The attributes are customer name, address, ID, and date last paid.

Each record requires a key field, or unique identifier. The **key field** in this file is the ID number; perhaps you'll use phone number because it will be unique for each record. Ask students to think about the pros and cons of using the telephone number (there may be more than one customer in the same house?) This is a very simplistic example of a database, but it should help students understand the terminology.

In a relational database, each table contains a **primary key**, a unique identifier for each record. To make sure the tables relate to each other, the primary key from one table is stored in a related table as a secondary key. For instance, in the customer table the primary key is the unique customer ID. That primary key is then stored in the order table as the secondary key so that the two tables have a direct relationship.

### Operations of a Relational DBMS

Relational database tables can be combined to deliver data required by users, provided that any two tables share a common **data element**.

Use these three basic operations to develop relational databases:

- **Select:** Create a subset of records meeting the stated criteria.
- **Join:** Combine related tables to provide more information than individual tables.
- **Project:** Create a new table from subsets of previous tables.

The biggest problem with these databases is the misconception that every data element should be stored in the same table. In fact, each data element should be analyzed in relation to other data elements, with the goal of making the tables as small in size as possible. The ideal relational database will have many small tables, not one big one. On the surface that may seem like extra work and effort, but by keeping the tables small, they can serve a wider audience because they are more flexible. This setup is especially helpful in reducing redundancy and increasing the usefulness of data.

### Hierarchical and Network DBMS

The hierarchical DBMS presents data to users in a treelike structure. Think of a mother and her children. A child only has one mother and inherits some of her characteristics, such as eye color or hair color. A mother might have one or more children to whom she passes some of her characteristics but usually not exact ones. The child then goes on to develop her own characteristics separate from the mother.

In a hierarchical database, characteristics from the parent are passed to the child by a pointer, just as a human mother will have a genetic connection to each human child. You can demonstrate this concept to students by showing them how this database pointer works by illustrating the simple hierarchy illustrated below.

A network data model is a variation of the hierarchical model. Take the same scenario with one parent and many children and add a father and perhaps a couple of stepparents. Now the parents aren't restricted to only one (the mother), but to many parents. That is, a parent can have many children and a child can have many parents. The parents pass on certain characteristics to the children, but the children also have their own distinct characteristics.



As with hierarchical structures, each relationship in a network database must have a pointer from all the parents to all the children and back, as this figure demonstrates.

These two types of databases are not easily manipulated and require extensive technical programming to meet changing requirements. Because they are difficult to build in the first place, some businesses are hesitant to replace them with newer relational data

models. They are referred to as legacy systems—systems that continue to be used because of the high cost of replacing them.

### **Object-Oriented Databases**

Many companies are moving away from strictly text-based database systems. Data as objects can be pictures, groups of text, voice, and audio. **Object-oriented databases** bring the various objects from many different sources and get them working together. If you combine the capabilities of a relational DBMS and an object-oriented database, you create an **object-relational DBMS**.

The next time you go to your dentist's office, you might see a good example of an object-oriented database management system. Many sophisticated dental database programs include a traditional text-based record of your treatment history, and will also include objects such as computer-stored X-ray films, and maybe a digital photograph of the inside of your mouth. All these objects are maintained as a database record. When you visit your dentist, she can retrieve your record on the computer terminal, update your treatment history, and take new X-rays and a new digital photo, all on the computer. On the screen, she can compare last year's X-rays with this year's. She may even use a graphic tooth chart to mark which teeth need attention.

## **CAPABILITIES OF DATABASE MANAGEMENT SYSTEMS**

A Database Management System (DBMS) is basically another software program like Word or Excel or e-mail. This type of software is more complicated; it permits an organization to centralize data, manage them efficiently, and provide access to the stored data by application programs. A DBMS has three components, all of them important for the long-term success of the system.

**Data definition language.** Marketing looks at customer addresses differently from Shipping, so you must make sure that all database users are speaking the same language. Think of it this way: marketing is speaking French, production is speaking German, and human resources is speaking Japanese. They are all saying the same thing, but it's very difficult for them to understand each other. Defining the data definition language itself sometimes gets shortchanged. The programmers who are creating the language sometimes say "Hey, an address is an address, so what." That's when it becomes critical to involve users in the development of the data definition language.

**Data dictionary.** Each data element or field should be carefully analyzed when the database is first built or as the elements are later added. Determine what each element will be used for, who will be the primary user, and how it fits into the overall scheme of things. Then write it all down and make it easily available to all users. This is one of the most important steps in creating a good database.

### **Querying and Reporting**



**Data manipulation language.** This is a formal language programmers use to manipulate the data in the database and make sure they are formulated into useful information. The goal of this language should be to make it easy for users. The basic idea is to establish a single data element that can serve multiple users in different departments, depending on the situation. Otherwise, you'll be employing programmers to get information from the database that users should be able to get on their own.

Data manipulation languages are getting easier to use and more prevalent. **SQL (Structured Query Language)** is the most prominent language and is now embedded in desktop applications such as Microsoft Access.

Because SQL is becoming a popular, easy method of extracting data, let's look at a couple of the commands it uses.

- **Select Statement:** Used to query data for specific information
- **Conditional Selection:** Used to specify which rows of a table are displayed, based on criteria contained in the WHERE clause
- **Joining Two Tables:** Used to combine data from two or more tables and display the results

SQL commands can be embedded in application programs written in many different languages. The manipulative characteristics of SQL have led to its popularity.

**Bottom Line: Database Management Systems have three critical components: the data definition language, the data manipulation language, and the data dictionary. Managers should ensure that all three receive attention. There are three types of databases: hierarchical, network, and relational. Relational databases are becoming the most popular of the three because they are easier to work with, easier to change, and can serve a wider range of needs throughout the organization. Managers should also make sure that end users are fully involved in developing organizational databases.**

## DESIGNING DATABASES

Before creating a database using a DBMS, students need to design the database. Emphasize that they should consider how the information is used, and how it is organized, and stored. Have them consider how this information could be organized better and used more easily throughout the organization.

### Normalization and Entity-Relationship Diagrams

Have the students determine the relationships between each data element that is currently used (**entity-relationship diagram**). The data don't necessarily have to be in a computer

for you to consider the impact. Determine which data elements work best together and how you will organize them in tables. Break your groups of data into as small a unit as possible (**normalization**). Even when you say it's as small as it can get, go back through again. Avoid redundancy between tables. Decide what the key identifier will be for each record. Emphasize that work at this stage will save money and time in the long run.

The example in the textbook is a good example to work through to show the normalization technique. You can also use examples of class registrations, hotel reservations, or video store rentals.

### **Distributing Databases**

A **distributed database**, which is stored in more than one physical location, is usually found in very large corporations that require immediate, fast access to data at multiple sites. There are two ways to structure distributed databases:

1. Partition a central database so that each remote processor has the necessary data to serve its local area.
2. Replicate the central database at all remote locations.

As the book points out, there are lots of disadvantages so you should be careful to determine if this is the right way for you to run your business.

## **6.3**

## **USING DATABASES TO IMPROVE BUSINESS PERFORMANCE AND DECISION MAKING**

---

Corporations and businesses go to great lengths to collect and store information on their suppliers and customers. What they haven't done a good job of in the past is fully using the data to take advantage of new products or markets. They're trying, though, as we see in this section.

### **DATA WAREHOUSES**

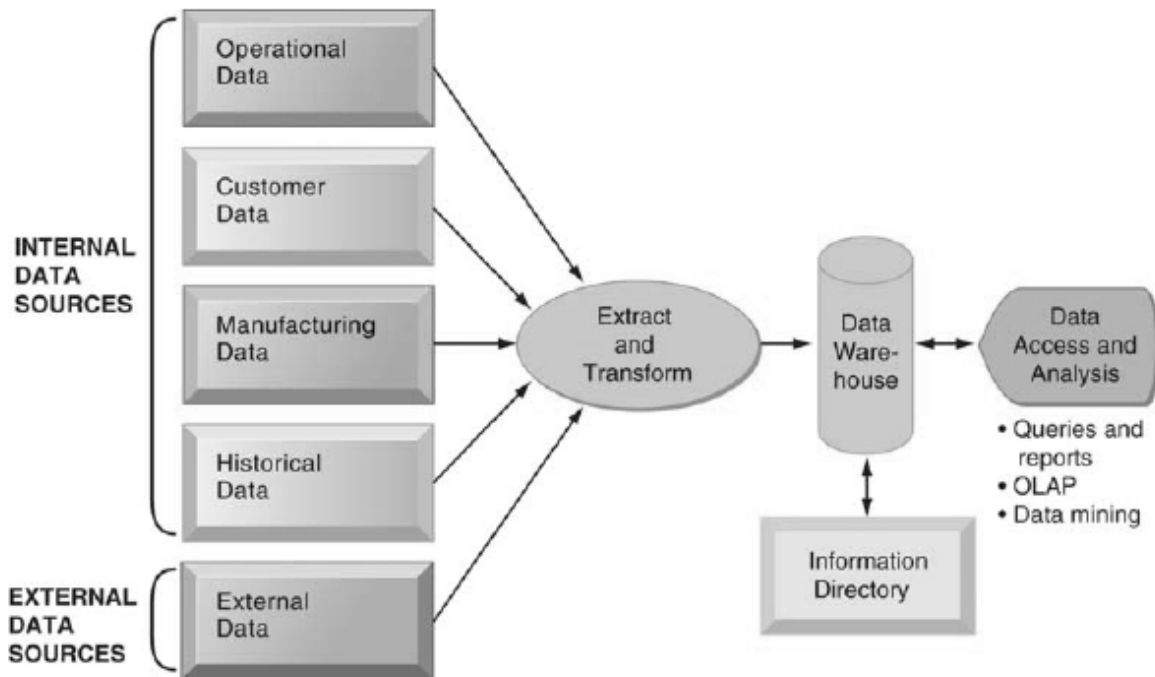
As organizations want and need more information about their company, their products, and their customers, the concept of data warehousing has become very popular. Data warehouses store old and new data about anything and everything that a company wants to maintain information on.

#### **What Is a Data Warehouse?**

A **data warehouse** is a database that stores current and historical data of potential interest to decision makers throughout the company. The data warehouse consolidates and standardizes information from different operational databases so that the information can be used across the enterprise for management analysis and decision making.

The data come from a variety of sources, both internal and external to the organization. They are then stored together in a data warehouse from which they can be accessed and analyzed to fit the user's needs.

**FIGURE 6-13 COMPONENTS OF A DATA WAREHOUSE**



## WINDOW ON MANAGEMENT: CANADIAN TIRE WAREHOUSES BUSINESS INTELLIGENCE FOR PROFITABILITY

### TO THINK ABOUT QUESTIONS

#### 1. Why was it so difficult for Canadian Tire to analyze the data it had collected?

Through mergers and acquisitions the company consisted of disparate units, whose data were stored separately. There was no enterprise-wide view of the data because units were separate and there was a lot of data to collect from the many stores. In addition to not having a consolidated and adequate data warehouse, there was also a lack of analytic software to explore the data that were collected.

#### 2. What kind of challenges did Canadian Tire encounter when implementing its data warehouse? What management, organization, and technology issues had to be addressed?

Management: It took a long time to hire a person with the requisite skills. Management would need to be supportive of the initiative (cost, time). They needed to have strong

oversight of the vendors, especially when there were two external suppliers to manage (they are now one).

Organization: Units were used to having separate ownership of data. They would need to understand and support integration of the data.

Technology: They were using an older mainframe; they would need to migrate to newer hardware and software. Other systems may be impacted. Users would need training. They might need a new network to handle the collection of data from distributed stores.

**3. How did the data warehouse improve decision making and operations at Canadian Tire? Are there benefits to customers?**

They could compare stores and find best practices, problems, or trends that could inform business decision making (ex. Effective product placements, sales trends, inventory management (using RFID, for example), and discovery of theft). They had real-time data that would allow faster response to issues. Customers would benefit if the improved decision making leads to better cost control: savings could be passed along to customers. Also, customers would benefit by better inventory management (items available in stores), sales that are targeted to certain locations, and improved customer service (based upon knowledge of demographics, trends, etc.).

**4. Do you think the issues facing the Bank of Montreal were the same or similar to those faced by Canadian Tire?**

Yes, some of the issues are the same: the data was silo-ed in branches and there was no enterprise-wide data warehouse. They would have had similar issues with looking at trends and operational effectiveness, although Canadian Tire would have had more issues with inventory management.

**MIS IN ACTION QUESTIONS**

**1. Go to [www.cognos.ca](http://www.cognos.ca) and find information about Cognos' business intelligence products. How much training do you think would be required to learn how to use a Cognos system?**

From the website:

***Your business benefits***

*Turn your investment in IBM® Cognos® 8 Business Intelligence into a strategic and operational advantage for your organization.*

- *Accelerate know-how across the organization, fully leveraging the functionality in your IBM Cognos 8 Business Intelligence solution.*
- *Get results by increasing the confidence of your users to take advantage of the software effectively.*
- *Empower users with the flexible training options to ramp them up quickly and efficiently for optimal performance, reducing the calls to your helpdesk.*

- *Ensure success by using certified IBM Cognos Education experts to empower your organization to drive maximum value from the IBM Cognos solution and benefit from time and cost savings in current and future projects.*

***Find the training that's right for you***

*Complete curricula have been designed to meet the diverse needs of every IBM Cognos 8 Business Intelligence user, including Administrator, Business Author, Professional Author, Consumer, Developer and Project Manager. For existing customers, we also offer "What's New" courses.*

- [IBM Cognos 8 BI Curricula](#)
- [Search and register for a course](#)

From this, we can see that there are different training needs depending on the user's role: an end user would need some training, but much of the output is presented as scorecards which are generally intuitive and easy to understand for most business end users. The website provides more detail on the courses/training available to contributors, administrators/modellers/project managers.

**2. How could Canadian Tire managers use their data to help them make decisions related to ordering, inventory, and pricing?**

Students can look at Wal-Mart to understand the full power of data mining. They can link the data warehouse to their inventory management system to allow automatic (just-in-time) ordering from suppliers. They could use data mining to look at cross selling opportunities (e.g. product placement decisions based on past purchases). They can mine the data to understand which items sell well in different geographic areas and have price discrimination based on this intelligence.

**Data Marts**

Since the data warehouse can be cumbersome, a company can break the information into smaller groups called **data marts**. It's easier and cheaper to sort through smaller groups of data. It's still useful to have a huge data warehouse, though, so that information is available to everyone who wants or needs it. You can let the user determine how the data will be manipulated and used. Using a data warehouse correctly can give management a tremendous amount of information that can be used to trim costs, reduce inventory, put products in the right stores at the right time, and attract new customers.

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

Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions are often referred to as **business intelligence (BI)**. Principal tools for business intelligence include software for database query and

reporting, tools for multidimensional data analysis (online analytical processing), and data mining.

### Online Analytical Processing (OLAP)

As technology improves, so does our ability to manipulate information maintained in databases. Have you ever played with a Rubik Cube — one of those little multicolored puzzle boxes you can twist around and around to come up with various color combinations? That's a close analogy to how multidimensional data analysis or **online analytical processing (OLAP)** works. In theory, it's easy to change data around to fit your needs.

### Data Mining

By using **data mining**, a digital firm can get more information than ever before from its data. One danger in data mining is the problem of getting information that on the surface may seem meaningful, but when put into context of the organization's needs, simply doesn't provide any useful information.

For instance, data mining can tell you that on a hot summer day in the middle of Ontario, more bottled water is sold in convenience stores than in grocery stores. That's information managers can use to make sure more stock is targeted to convenience stores. Data mining could also reveal that when customers purchase white socks, they also purchase bottled water 62 percent of the time. We seriously doubt there is any correlation between the two purchases. The point is that you need to beware of using data mining as a sole source of decision making and make sure your requests are as focused as possible.

Many companies collect lots of data about their business and customers. The most difficult part has been to turn that data into useful information. With improved database technology, organizations are creating new opportunities for connecting with their customers by extracting information easier and more precisely from their data warehouses. Firms are using better data mining techniques to target customers and suppliers with just the right information at the right time.

The types of information obtainable from data mining include:

- *Associations* — occurrences linked to a single event.
- *Sequences* — events are linked over 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.
- *Clustering* — works in a manner similar to classification when no groups have yet been defined.
- *Forecasting* — uses a series of existing values to forecast what other values will be.

**Predictive analysis** uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events, such as the probability a customer will respond to an offer or purchase a specific product.

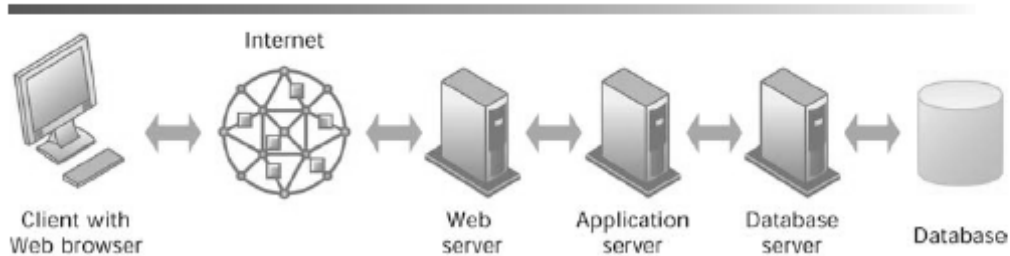
## DATABASES AND THE WEB

Even though Web browsers have only been around for a few years, they are far easier to use than most of the query languages associated with the other programs on mainframe computer systems. Many companies are finding out that it's easier to provide their "road warriors" with Web-like browsers attached to the computer at the main office. Employees anywhere can have up-to-the-minute access to any information they need. It's also proving cheaper to create "front-end" browser applications that can more easily link information from disparate systems than to try to combine all the systems on the "back-end".

One of the easiest ways to make databases available to any user is by linking the internal databases to the Web through software programs that provide a connection to the database without major reconfigurations. A database server, which is a special dedicated computer, maintains the DBMS. A software program, called an application server, processes the transactions and offers data access. A user making an inquiry through the Web server can connect to the organization's database and receive information in the form of a Web page.

Figure 6-16 shows how an application server provides the interface between the database and the Web.

**FIGURE 6-16 LINKING INTERNAL DATABASES TO THE WEB**



The benefits of using a Web browser to access a database are as follows:

- Ease-of-use
- Less training for users
- No changes to the internal database
- Cheaper than building a new system
- Creating new efficiencies and opportunities
- Provide employees with integrated firmwide views of information

**Bottom Line:** There are many ways to manipulate databases so that an organization can save money and still have useful information. With technological improvements companies don't have to continually start from scratch but can blend the old with

**the new when they want to update their systems. The Web is the perfect delivery vehicle for databases and is cheaper than building proprietary systems.**

## **WINDOW ON TECHNOLOGY: THE DATABASES BEHIND MYSPACE**

### **TO THINK ABOUT QUESTIONS**

#### **1. What kind of databases and database servers does MySpace use?**

In its initial phases, MySpace operated with two Web servers communicating with one database server and a Microsoft SQL Server database. The site continued adding Web servers to handle increased user requests. After the number of accounts exceeded 500,000 the site added more SQL Server databases: one served as a master database, the others focused on retrieving data for user page requests. After two million accounts were activated, MySpace switched to a vertical partitioning model in which separate databases supported distinct functions of the Web site. After three million accounts, the site scaled out by adding many cheaper servers to share the database workload.

It eventually switched to a virtualized storage architecture in which databases write data to any available disk, thus eliminating the possibility of an application's dedicated disk becoming overloaded. MySpace later installed a layer of servers between the database servers and the Web servers to store and serve copies of frequently accessed data objects so that the site's Web servers wouldn't have to query the database servers with lookups as frequently.

#### **2. Why is database technology so important for a business such as MySpace?**

Almost everything MySpace receives from and serves to its users are data objects like pictures, audio files and video files. The objects are very individualized and attached to a certain entity (person). Its databases must make the objects readily available to anyone requesting access to that entity. Database technology is the only technology that accomplishes the mission.

#### **3. How effectively does MySpace organize and store the data on its site?**

In its infancy, MySpace used two Web servers communicating with one database server. That was adequate when the site had a small number of users who were updating or accessing database objects. Obviously that won't work with tens of millions of users. Unfortunately, MySpace still overloads more frequently than other major Web sites. With a log-in error rate of 20 to 40 percent on some days, the site is not effectively organizing or storing data at all.

#### **4. What data management problems have arisen? How has MySpace solved, or attempted to solve, these problems?**



Some of the problems MySpace has encountered are inadequate storage space on its database servers, slow access or no access through its log-in application, and users' inabilities to access data. Over the years, MySpace has attempted to fix these problems by adding more Web servers and more database servers. Some were simply "added on" without restructuring the entire system to more efficiently use its hardware and software. Workloads were not distributed evenly between servers which caused inefficient use of resources. MySpace developers continue to redesign the Web site's database, software, and storage systems, to keep pace with its exploding growth, but their job is never done.

## MIS IN ACTION QUESTIONS

**Explore MySpace.com, examining the features and tools that are not restricted to registered members. Then answer the following questions:**

**1. Based on what you can view without registering, what are the entities in MySpace's database?**

Obviously, individual users are the main entity in MySpace's databases. Other entities are video files, audio files, blogs, forums, groups, events, favourites, and email.

**2. Which of these entities have some relationship to individual members?**

Which of the entities have a relationship to individual members depends on what the individual decides. For instance, it's possible that Sarah would have a list of films (video files) attached to her profile. She may also participate in forums or groups. It's possible that all the entities have some relationship to individual members.

**3. Select one of these entities and describe the attributes for that entity.**

Films included in MySpace's databases likely have these attributes: name, date produced, date released, actors, actresses, director, subject, place it was filmed, musical scores included in the film, awards given to the film, comments of film goers, and critics' ratings.

## 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 organization needs to have rules on how the data are to be organized and maintained, and who is allowed to view the data or change them.

**Information policy** — No one part of the organization should feel that it owns information to the exclusivity of other departments or people in the organization. A certain department may have the primary responsibility for updating and maintaining the information, but that department still has to share it across the whole company. Well-written information policies can outline the rules for using this important resource, including how it will be shared, disseminating, acquiring, standardizing, classifying, and inventorying information.

**Data administration** is responsible for the specific policies and procedures through which data can be managed as an organizational resource. This function can help define and structure the information requirements for the entire organization to ensure it receives the attention it deserves. Data administration is responsible for the following:

- Developing information policies
- Planning for data
- Overseeing logical database design
- Data dictionary development
- Monitoring the usage of data by techies and non-techies

**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.

**Bottom Line: Information is power. The more information users have in an easy-to-use form, the more they can accomplish. Managers need to consider information as an important resource for which everyone has a responsibility.**

## ENSURING DATA QUALITY

Poor data quality can have far-reaching implications that can make a company legally liable. Although all employees may share in the responsibility for maintaining good quality data, managers obviously have the greater share. And, there is a lot more to a useful database than simply listing a bunch of data elements and hoping people use them as intended.

**Data quality audits** can help companies identify the accuracy and completeness of data. **Data cleansing** also known as *data scrubbing* can detect and correct data and enforce consistency among different sets of data. This last point is important if an organization has combined several databases from different sources since chances are great that there are erroneous or mismatched data.

**Bottom Line: As with any other resource, managers must administer their data, plan their uses, and discover new opportunities for the data to serve the organization through changing technologies.**

## 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?***

A computer system organizes data in a hierarchy that starts with bits and bytes and progresses to fields, records, files, and databases. 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.

**2. *What are the major capabilities of a database management system (DBMS), and why is a relational DBMS so powerful?***

A database management system (DBMS) consists of software that permits centralization of data and data management so that businesses have a single consistent source for all their data needs. A single database services multiple applications. The most important feature of the DBMS is its ability to separate the logical and physical views of data. The user works with a logical view of data. The DBMS retrieves information so that the user does not have to be concerned with its physical location.

The principal capabilities of a DBMS include a data definition capability, a data dictionary capability, and a data manipulation language. The data definition language 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 is the primary method for organizing and maintaining data today in information systems. It organizes data in two-dimensional tables with rows and columns called relations. Each table contains data about an entity and its attributes. Each row represents a record and each column represents an attribute or

field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation.

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

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. An entity-relationship diagram graphically depicts the relationship between entities (tables) in a relational database. Database design also considers whether a complete database or portions of the database can be distributed to more than one location to increase responsiveness and reduce vulnerability and costs. There are two major types of distributed databases: replicated databases and partitioned databases.

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

Powerful tools are available to analyze and access the information in databases. A data warehouse consolidates current and historical data from many different operational systems in a central database for reporting and analysis. Data warehouses support multidimensional data analysis, also known as online analytical processing (OLAP). OLAP represents relationships among data as a multidimensional structure, which can be visualized as cubes of data and cubes within cubes of data, 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 further behaviour and guide decision making. 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

The following alphabetical list identifies the key terms discussed in this chapter.

**Attributes** — a piece of information describing a particular entity.

**Business Intelligence** — Applications and technologies to help users make better business decisions.

**Data administration** — a special organizational function for managing the organization's data resources that is concerned with information policy, data planning, maintenance of data dictionaries, and data quality standards.

**Data cleansing** — activities for detecting and correcting data in a database or file that are incorrect, incomplete, improperly formatted, or redundant. Also known as data scrubbing.

**Data definition language** — the component of a database management system that defines each data element as it appears in the database.

**Data dictionary** — an automated or manual tool for storing and organizing information about the data maintained in a database.

**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.

**Data inconsistency** — the presence of different values for the same attribute when the same data are stored in multiple locations.

**Data manipulation language** — a language associated with a database management system that end users and programmers use to manipulate data in the database.

**Data mart** — a small data warehouse containing only a portion of the organization's data for a specified function or population of users.

**Data mining** — analysis of large pools of data to find patterns and rules that can be used to guide decision making and predict future behaviour.

**Data quality audit** — a survey and/or sample of files to determine accuracy and completeness of data in an information system.

**Data redundancy** — the presence of duplicate data in multiple data files.

**Data warehouse** — a database with reporting and query tools that stores current and historical data extracted from various operational systems and consolidated for management reporting and analysis.

**Database** — a group of related files.

**Database (rigorous definition)** — a collection of data organized to service many applications at the same time by storing and managing data so that they appear to be in one location.

**Database administration** — refers to the more technical and operational aspects of managing data, including physical database design and maintenance.

**Database management system (DBMS)** — special software to create and maintain a database and enable individual business applications to extract the data they need without having to create separate files or data definitions in their computer programs.

**Database server** — a computer in a client/server environment that is responsible for running a database management system (DBMS) to process structured query language (SQL) statements and perform database management tasks.

**Distributed database** — a database that is stored in more than one physical location. Parts or copies of the database are physically stored in one location, and other parts or copies are stored and maintained in other locations.

**Entity** — a person, place, thing, or event about which information must be kept.

**Entity-relationship diagram** — a methodology for documenting databases illustrating the relationship between various entities in the database.

**Field** — a group of characters into a word, a group of words, or a complete number, such as a person's name or age.

**File** — A group of records of the same type.

**Foreign key** — field in a database table that enables users to find related information in another database table.

**Information policy** — formal rules governing the maintenance, distribution, and use of information in an organization.

**Key field** — a field in a record that uniquely identifies instances of that record so that it can be retrieved, updated, or sorted.

**Normalization** — the process of creating small stable data structures for complex groups of data when designing a relational database.

**Object-oriented DBMS** — an approach to data management that stores both data and the procedures acting on the data as objects that can be automatically retrieved and shared; the objects can contain multimedia.

**Object-relational DBMS** — a database management system that combines the capabilities of a relational database management system (DBMS) for storing traditional information and the capabilities of an object-oriented DBMS for storing graphics and multimedia.

**Online analytical processing (OLAP)** — capability for manipulating and analyzing large volumes of data from multiple perspectives.

**Predictive analysis** — Use of data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events.

**Primary key** — unique identifier for all the information in any row of a database table.

**Program-data dependence** — the close relationships between data stored in files and the software programs that update and maintain those files. Any change in data organization or format requires a change in all the programs associated with those files.

**Record** — a group of related fields.

**Relational DBMS** — a type of logical database model that treats data as if they were stored in two-dimensional tables. It can related data stored in one table to data in another as long as the two tables share a common data element.

**Structured query language (SQL)** — the standard data manipulation language for relational database management systems.

**Tuple** — a row or record in a relational database.

## REVIEW QUESTIONS

2. 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.

Figure 6–1 shows the data hierarchy. The data hierarchy includes bits, bytes, fields, records, files, and databases. Data are organized in a hierarchy that starts with the bit, which is represented by either a 0 (off) or a 1 (on). Bits can be grouped to form a byte to represent one character, number, or symbol. Bytes can be grouped to form a field, such as a name or date, and related fields can be grouped to form a record. Related records can be collected to form files, and related files can be organized into a database.

**Define and explain the significance of entities, attributes, and key fields.**

- Entity is a person, place, thing, or event on which information can be obtained.
- Attribute is a piece of information describing a particular entity.
- Key field is a field in a record that uniquely identifies instances of that unique record so that it can be retrieved, updated, or sorted. For example, a person's name cannot be a key because there can be another person with the same name, whereas a social security number is unique. Also a product name may not be unique but a product number can be designed to be unique.

**List and describe the problems of the traditional file environment.**

Problems with the traditional file environment include data redundancy and confusion, program-data dependence, lack of flexibility, poor security, and lack of data sharing and availability. Data redundancy is the presence of duplicate data in multiple data files. In this situation, confusion results because the data can have different meanings in different files. Program-data dependence is the tight relationship between data stored in files and the specific programs required to update and maintain those files. This dependency is very inefficient, resulting in the need to make changes in many programs when a common piece of data, such as the zip code size, changes. Lack of flexibility refers to the fact that it is very difficult to create new reports from data when needed. Ad-hoc reports are impossible to generate; a new report could require several weeks of work by more than one programmer and the creation of intermediate files to combine data from disparate files. Poor security results from the lack of control over the data because the data are so widespread. Data sharing is virtually impossible because it is distributed in so many different files around the organization.

**Define a database and a database management system and describe how it solves the problems of a traditional file environment.**

A database is a collection of data organized to service many applications efficiently by storing and managing data so that they appear to be in one location. It also minimizes redundant data. A database management system (DBMS) is special software that permits an organization to centralize data, manage them efficiently, and provide access to the stored data by application programs.

A DBMS can reduce the complexity of the information systems environment, reduce



data redundancy and inconsistency, eliminate data confusion, create program-data independence, reduce program development and maintenance costs, enhance flexibility, enable the ad hoc retrieval of information, improve access and availability of information, and allow for the centralized management of data, their use, and security.

## **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.**

A DBMS includes capabilities and tools for organizing, managing, and accessing the data in the database. The principal capabilities of a DBMS include data definition language, data dictionary, and data manipulation language.

- The data definition language 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.

### **Define a relational DBMS and explain how it organizes data.**

The relational database is the primary method for organizing and maintaining data today in information systems. It organizes data in two-dimensional tables with rows and columns called relations. Each table contains data about an entity and its attributes. Each row represents a record and each column represents an attribute or field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation.

### **List and describe the three operations of a relational DBMS.**

In a relational database, three basic operations are used to develop useful sets of data: select, project, and join.

- Select operation creates a subset consisting of all records in the file that meet stated criteria. In other words, select creates a subset of rows that meet certain criteria.
- Join operation combines relational tables to provide the user with more information that is available in individual tables.
- Project operation creates a subset consisting of columns in a table, permitting the user to create new tables that contain only the information required.

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

**Define and describe normalization and referential integrity and explain how they contribute to a well-designed relational database.**

Normalization is the process of creating small stable data structures from complex groups of data when designing a relational database. Normalization streamlines relational database design by removing redundant data such as repeating data groups. A well-designed relational database will be organized around the information needs of the business and will probably be in some normalized form. A database that is not normalized will have problems with insertion, deletion, and modification.

Referential integrity rules 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.

**Define a distributed database and describe the two main ways of distributing data.**

A distributed database is one that is stored in more than one physical location. A distributed database can be partitioned or replicated. When partitioned, the database is divided, so that each remote processor has access to the data that it needs to serve its local area. These databases can be updated locally and later justified with the central database. With replication, the database is duplicated at various remote locations. Figure 6–12 shows how a database is distributed.

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

**Define a data warehouse, explaining how it works and how it benefits organizations.**

A data warehouse is a database with archival, querying, and data exploration tools (i.e., statistical tools) and is used for storing historical and current data of potential interest to managers throughout the organization and from external sources (e.g., competitor sales or market share). The data originate in many of the operational areas and are copied into the data warehouse as often as needed. The data in the warehouse are organized according to company-wide standards so that they can be used for management analysis and decision making. Data warehouses support looking at the data of the organization through many views or directions. The data warehouse makes the data available to 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. The data warehouse system allows managers to look at products by customer, by year, by salesperson, essentially different slices of the data. Normal operational databases do not permit such different views.

**Define business intelligence and explain how it is related to database technology.**

Powerful tools are available to analyze and access information that has been captured and organized in data warehouses and data marts. These tools enable users to analyze the data to see new patterns, relationships, and insights that are useful for guiding decision making. These tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions are often referred to as business intelligence. Principal tools for business intelligence include software for database query and reporting tools for multidimensional data analysis and data mining.

**Describe the capabilities of online analytical processing (OLAP).**

Data warehouses support multidimensional data analysis, also known as online analytical processing (OLAP), which enables users to view the same data in different ways using multiple dimensions. Each aspect of information represents a different dimension.

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. OLAP enables users to obtain online answers to ad hoc questions in a fairly rapid amount of time, even when the data are stored in very large databases. Online analytical processing and data mining enable the manipulation and analysis of large volumes of data from many perspectives, for example, sales by item, by department, by store, by region, in order to find patterns in the data. Such patterns are difficult to find with normal database methods, which is why a data warehouse and data mining are usually parts of OLAP. 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.

**Define data mining, describing how it differs from OLAP and the types of information it provides.**

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

**Explain how text mining and Web mining differ from conventional data mining.**

Conventional data mining focuses on data that have been structured in databases and files. Text mining concentrates on finding patterns and trends in unstructured data contained in text files. The data may be in email, memos, call center transcripts,

survey responses, legal cases, patent descriptions, and service reports. Text mining tools extract key elements from large unstructured data sets, discover patterns and relationships, and summarize the information.

Web mining helps businesses understand customer behaviour, evaluate the effectiveness of a particular Web site, or quantify the success of a marketing campaign. Web mining looks for patterns in data through

- Web content mining: extracting knowledge from the content of Web pages
- Web structure mining: examining data related to the structure of a particular Web site
- Web usage mining: examining user interaction data recorded by a Web server whenever requests for a Web site's resources are received

**Describe how users can access information from a company's internal databases through the Web.**

Conventional databases can be linked via middleware to the Web or a Web interface to facilitate user access to an organization's internal data. Web browser software on his/her client PC is used to access a corporate Web site over the Internet. The 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 special middleware software that then translates HTML commands into SQL so that they can be processed by the DBMS working with the database. The DBMS receives the SQL requests and provides the required data. The 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. The software working between the Web server and the DBMS can be an application server, a custom program, or a series of software scripts.

**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.**

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.

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.

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.

**Explain why data quality audits and data cleansing are essential.**

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.

A data quality audit 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 consists of activities for detecting and correcting data in a database that are incorrect, incomplete, improperly formatted, or redundant. Data cleansing not only corrects data but also enforces consistency among different sets of data that originated in separate information systems.

## **Discussion Questions**

**1. It has been said that you do not need database management software to create a database environment. Discuss.**

A database is a collection of data organized to service many applications at the same time by storing and managing data so that they appear to be in one location. It is not mandated that a database have a DBMS. What is most important is the concept of a database — a model for organizing information so that it can be stored and accessed flexibly and efficiently. Without the right vision of a database and data model, a DBMS is not effective. A DBMS is special software to create and maintain a database. It enables individual business applications to extract the data they need without having to create separate files or data definitions in their computer programs. However, the use of a DBMS can reduce program-data dependence along with program development and maintenance costs. Access and availability of information can be increased because users and programmers can perform ad-hoc queries of data in the database. The DBMS allows the organization to centrally manage data, its use, and security.

**2. To what extent should end users be involved in the selection of a database**

**management system and database design?**

End users should be involved in the selection of a database management system and the database design. Developing a database environment requires much more than just selecting the technology. It requires a change in the corporation's attitude toward information. The organization must develop a data administration function and a data planning methodology. The end-user involvement can be instrumental in mitigating the political resistance organizations may have to many key database concepts, especially to sharing information that has been controlled exclusively by one organizational group.

## **COLLABORATION AND TEAMWORK: IDENTIFYING ENTITIES AND ATTRIBUTES IN AN ONLINE DATABASE**

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

Direct your students to the Web sites below. In their analysis, students should quickly articulate that many of these sites use the same entities and attributes to keep track of their database.

There are hundreds of Internet Movie Databases so students will have to select the one that interests them. The Web sites for AOL Music and iGo.com are listed below.

<http://music.aol.com/>

<http://igo.com/>

## **LEARNING TRACK MODULES**

*Database Design, Normalization, and Entity-Relationship Diagramming*  
*Introduction to SQL*  
*Hierarchical and Network Data Models*

Students will find Learning Track Modules on these topics at the MyMISLab for this chapter.

## **HANDS-ON MIS: PROJECTS**

## Management Decision Problems

**1. Emerson Process Management:** data warehouse was full of inaccurate and redundant data gathered from numerous transaction processing systems. The design team assumed all users would enter data the same way. Users actually entered data in multiple ways. Assess the potential business impact of these data quality problems. What decisions have to be made and steps taken to reach a solution?

Managers and employees can't make accurate and timely decisions about customer activity because of inaccurate and redundant data. The company could be wasting resources pursuing customers it shouldn't and neglecting its best customers. The company could be experiencing financial losses resulting from the inaccurate data.

Managers, employees, and data administrators need to identify and correct the faulty data and then establish better routines for editing data when it's entered. The company should perform a data quality audit by surveying entire data files, surveying samples from data files, or surveying end users for perceptions of data quality. The company needs to perform data cleansing operations to correct errors and enforce consistency among the different sets of data at their origin.

**2. Industrial supply company:** the company wants to create a single data warehouse by combining several different systems. The sample files from the two systems that would supply the data for the data warehouse contain different data sets.

1. What business problems are created by not having these data in a single standard format?

Managers are unable to make good decisions about the company's sales and products because of inconsistent data. Managers can't determine which products are selling the best world-wide; they can only determine product sales by region.

2. 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.

It may not be too hard to create a database with a single standard format if the company used middleware to pull both data sets into the consolidated database. The company should use specialized data-cleansing software that would automatically survey data files, correct errors in the data, and integrate the data in a consistent company-wide format. Problems that may occur would stem from inconsistent data names like the Territory and Customer ID in the old sets and data element names like Division in the new set. The data administrators, managers, and employees may have to track the data conversion and manually convert some data.

3. Should the problems be solved by database specialist or general business managers? Explain.

Both the database specialist and general business managers should help solve the problems. Data administrators are responsible for developing information policy, planning for data, overseeing logical database design and data dictionary development, and monitoring how information system specialists and end-user groups use data. However, end-users and business managers have the final decision-making authority and responsibility for the data.

4. Who should have the authority to finalize a single company-wide format for this information in the data warehouse?

Owners and managers are the only ones who have the authority to finalize the format for the information in the data warehouse. They could develop an information policy that specifies the organization's rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying information.

## **ACHIEVING OPERATIONAL EXCELLENCE. BUILDING A RELATIONAL DATABASE FOR INVENTORY MANAGEMENT**

**Software skills: Database design, querying and reporting**

**Business skills: Inventory management**

This exercise requires that students know how to create queries and reports using information from multiple tables. The solutions provided here were created using the query wizard and report wizard capabilities of Access. Students can, of course, create more sophisticated reports if they wish.

The data for students for this exercise is found in the file named **Inventory\_Mgmt\_Q.mdb** in the Chapter 6 folder.

The database would need some modification to answer other important questions about the business. The owners might want to know, for example, which are the fastest-selling bicycles. The existing database shows products in inventory and their suppliers. The owners might want to add an additional table (or tables) in the database to house information about product sales, such as the product identification number, date placed in inventory, date of sale, purchase price, and customer name, address, and telephone number. Management could use this enhanced database to create reports on best selling bikes over a specific period, the number of bicycles sold during a specific period, total volume of sales over a specific period, or best customers. Students should be encouraged to think creatively about what other pieces of information should be captured on the database that would help the owners manage the business.

1. 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.



2. Prepare a report that lists each supplier, its products, the quantities on hand, and associated reorder levels. The report should be sorted alphabetically by supplier. Within each supplier category, the products should be sorted alphabetically.
3. 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.
4. 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?

The answers to the following questions can be found in the file named **Inventory\_Mgmt\_S.mdb in the Chapter 6 folder.**

## **IMPROVING DECISION MAKING: SEARCHING ONLINE DATABASES FOR OVERSEAS BUSINESS RESOURCES**

**Software skills: Online database**

**Business Skills: Researching services for overseas operations**

The Internet is a valuable source of databases where users can search for services and products in areas or countries that are far from their own locations. Your company is located in Calgary, Alberta and manufactures office furniture of various types. You have recently acquired several new customers in Australia, and a study you commissioned indicates that with a presence there, you could greatly increase your sales. Moreover, your study indicates that you could do even better if you actually manufactured many of your products locally (in Australia). First you need to set up an office in Melbourne to establish a presence, and then you need to begin importing from Canada. You then can plan to start producing locally.

You will soon be traveling to the area to make plans to actually set up an office, and you want to meet with organizations that can help you with your operation. You will need to engage people or organizations that offer many services necessary for you to open your office, including lawyers, accountants, import-export experts, telecommunications equipment and support, and even trainers who can help you to prepare your future employees to work for you.

1. **List the companies you would contact to interview on your trip to determine whether they can help you with functions you think are vital to establishing your office.**

Start by searching for Canadian government Department of Foreign and International Trade (DFAIT)

([http://www.international.gc.ca/commerce/index.aspx?menu\\_id=12&menu=L](http://www.international.gc.ca/commerce/index.aspx?menu_id=12&menu=L)) for advice on doing business in Australia. Information on how to obtain permits, and legislation around doing business internationally can be found on this site. You may also find information on trade missions. You can also check your provincial

government website for information on any trade missions to Australia, although most provincial missions are coordinated by the DFAIT office.

Then try the following online databases to locate companies that you would like to meet with during your coming trip: Australian Business Register, Australia Trade Now, and the Nationwide Business Directory of Australia. If necessary, you should also try search engines such as Yahoo! ([www.yahoo.com](http://www.yahoo.com)) and Google ([www.google.com](http://www.google.com)).

2. **Rate the databases you used for accuracy of name, completeness, ease-of-use, and general helpfulness.**
3. **What does this exercise tell you about the design of databases?**

Students may not understand that the World Wide Web is one massive data warehouse, but in non-technical terms that is exactly what it is. Remind them of this when they are completing this assignment. This assignment may best be accomplished in groups, where they can consolidate their findings into a written or oral presentation.

## **CASE STUDY: THE RCMP AND ITS DATA: MORE AND MORE AND MORE DATA – AND WHAT THE RCMP DOES WITH ITS DATA?**

1. **Evaluate the RCMP's reasons for implementing STaCS and CADVIEW. Compare the need for efficiency with the need for effectiveness. Were both efficiency and effectiveness key factors in these solutions?**

The systems are used to prevent, solve, track, and report crimes. STaCS allows the process by which DNA is evaluated to be viewed by those involved in the case. A DNA sample can be tracked and monitored, to ensure the process is not compromised. CADVIEW allows law enforcement to see where criminal activity is occurring. This can be used for scheduling resources, and for evaluating and responding to trends.

Students can also refer to earlier Window on Technology to discuss the use of the DNA database to solve crimes, and to prevent the innocent from being convicted.

In this case, effectiveness is probably more important than efficiency, although in some cases efficiency would be more critical. For example, in most cases it is more important to have a case solved (or prevented) rather than doing it in the least time or with the least amount of resources. There may be cases (such as kidnapping) where efficiency (time) is more important, and these systems would support that by providing lists of people with matching DNA. Overall, efficiency was more important for the processing of the DNA samples (STaCS) and effectiveness was key for the

CADVIEW system (allocating resources, preventing crime). However, STaCS does provide increases in productivity and therefore cost savings.

**2. What management, organization, and information technology factors were responsible for the decision to implement these systems?**

Management: recognition that there is a need to have a centralized database, and allow sharing of data between jurisdictions; recognition of potential for increased efficiency; willingness to support; understanding of culture

Organization: highly organized and structured; top down, so implementation and buy-in to new system was “seamless.”

Technology: Software available to allow them to create these systems

**3. What problems were to be addressed by these systems? How did data management address these problems? How effective are these solutions?**

- The need to have a centralized databank of DNA records to allow sharing cross-jurisdictions
- To prevent crimes
- To solve crimes
- To prevent improper incarceration
- To have better information on crime activity (rates, types, locations) with a view to prevention
- Better allocation of resources
- Ensure proper processing of DNA samples

**4. What are the privacy issues that arise from the use of these two databases? Are there simple solutions to these issues, or are the privacy issues complex, requiring the courts to adjudicate them?**

These are complex issues, and include ethical issues. Citizens resist having this data stored about them: it is vulnerable to hackers, it may be used for purposes undefined, and it may not be accurate. The courts will be called in to offer opinions and judgments as situations arise as these are new areas. There is little legislation that governs the use of these databases. There is an obligation for the management to ensure the data is secure, but the consequences and penalties for non compliance will need to be established by the courts.