1

INTRODUCING DATA AND DATA MANAGEMENT

Starting Point

Go to www.wiley.com/college/gillenson to assess your knowledge of data and data management.
Determine where you need to concentrate your effort.

What You'll Learn in This Chapter

- The role of data and data management as a business resource
- Identifying business processes and related data
- Potential concerns about data management

After Studying This Chapter, You'll Be Able To

- Understand the role of data and databases as a business resource and be able to use data to gain a competitive advantage
- Evaluate your organization's core business and primary processes to identify data collection requirements and sources in order to develop a data management system for your organization
- Identify potential human data sources and the data you can expect to retrieve based on organizational roles
- Identify key data management concerns in your organization and understand how to protect your organization's data from loss or corruption

INTRODUCTION

We live in a true information age where data is a critical resource. Today, perhaps more than ever before, knowledge is power. This chapter provides an introduction to data and some fundamental points about data management. We start with the role that data has played and continues to play as a business resource. We then look at how business practices help you understand data collection requirements. Finally, we take a brief look at some potential data management concerns.

1.1 Understanding the Role of Data and Databases

It may sometimes seem to you as if every aspect of your life is categorized and tracked from birth to death. In addition, you have direct access to more information than you could ever possibly use through the Internet.

We are being drowned in a veritable sea of data. Much of it is potentially valuable, but the situation has reached the point where data is being gathered so fast that much of it may never be put to use. Although the terms *data* and *information* are often used interchangeably, it is important to understand the distinction between them. **Data** (sometimes called "raw data") is a stream of facts that is not organized or arranged into a form that people can understand or use. Raw data has very little value. For instance, a retail store register captures a variety of raw data, such as the barcode associated with a purchased item, its price, the date and time, and so on. Each of these pieces of data has little meaning in isolation. However, when that raw data is organized by combining it with other similar data and is analyzed, meaningful information, such as the total number of sweaters sold at a particular store on a particular day, can be generated. **Information** is data that has been organized in such a way as to be meaningful and useful to people. One way to convert raw data into information is to use a database. A **database** is an ordered collection of related data elements intended to meet the information needs of an organization and designed to be shared by multiple users. Databases are our most powerful tool for organizing data into potentially valuable information.

Data is indispensable to every kind of modern business and government organization. Data, the applications that process the data, and the computers on which the applications run are fundamental to every aspect of every kind of endeavor. When speaking of corporate resources, people used to list such items as capital, plant and equipment, inventory, personnel, and patents. Today, any such list of corporate resources must include the corporation's data. It has even been suggested that data is the most important corporate resource because it describes all of the others.

Data can give a company a crucial competitive advantage. We routinely speak of data and the information derived from it as competitive weapons in hotly contested industries. For example, FedEx had a significant competitive advantage when it first provided access to its package-tracking data on its Web site. Then, once one company in an industry develops a new application that takes advantage of its data, the other companies in the industry are forced to match it to remain competitive.

This cycle continually moves the use of data to ever higher levels, making it an even more important corporate resource. Examples of this abound. Banks provide their customers with online access to their accounts. Package shipping companies provide up-to-the-minute information on the whereabouts of a package. Retailers send manufacturers product sales data that the manufacturers use to adjust inventories and production cycles. Manufacturers automatically send their parts suppliers inventory data and expect the suppliers to use the data to keep a steady stream of parts flowing.

1.1.1 A Practical Example

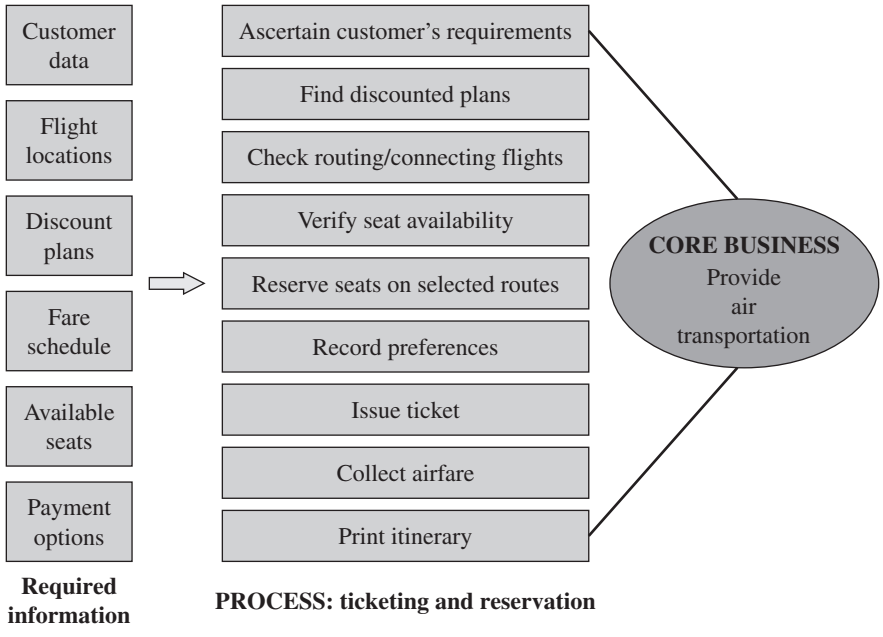
Let's take a look at a practical example. Airline companies need an overwhelming amount of data to survive. This includes internal data about their own resources and personnel, but also data about their customers' travel habits and factors that influence their business such as changes in fuel costs or actions taken by competitors.

To understand this, take a look at one small part of the airline business, the process of ticketing and seat reservation. Examine the individual tasks involved in the ticketing and seat reservation process. Let's say that a customer wants to make a trip from New York to Miami on a Monday. Are there any promotional discount plans? The ticketing agent must find and offer a suitable discount plan, if available. Although the agent has the desk, equipment, and other resources, he or she needs information about the available discount plans.

The next task involves checking possible routing options and connecting flights. To perform this task, the agent needs information about the different routing options and connecting flights on Mondays. After going through the routing options and flight times, let's say that the customer has picked a specific routing option. The next task in the process is verifying seat availability. What information does the agent need to complete this task? The agent must have information about seat availability in each leg of the journey. The agent needs different types of information for every task of the process. Figure 1-1 shows information needed for the ticketing and seat reservation process. Notice the various types of information necessary for the process.

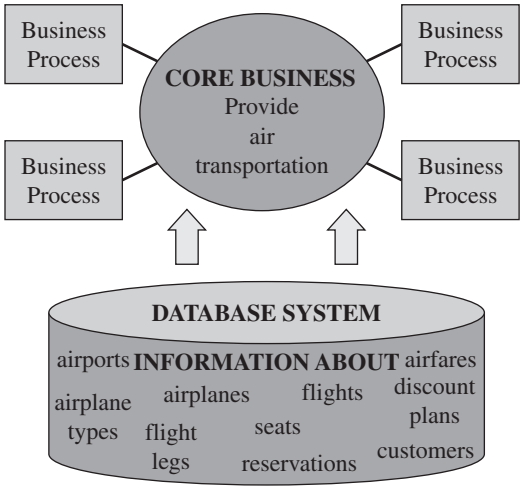
We have considered just one process in the airline company. Even for this one process, various types of critical information are essential. Now, consider the number of additional processes involved in day-to-day business activities, each process with its own critical data set (see Figure 1-2). From this, you can start

Figure 1-1



Information needed for ticketing and seat reservation.

Figure 1-2



Database system needed to support airline company core business activities.

to get an idea of the volume of information needed by any business or other organization.

1.1.2 Understanding Data Management

Data is a difficult corporate resource to manage. In data, you have a resource of tremendous volume, with billions, trillions, and more individual pieces, each piece of which is different from the next. And much of it is in a state of change at any one time.

As far back as the early to mid-1960s, companies began to realize that storing each application's data separately, in simple files, was problematic for the following reasons:

- The increasing volume of data.
- The increasing demand for data access.
- The need for data security, privacy, backup, and recovery.
- The desire to share data and cut down on data redundancy (unwanted duplicate data in a database).

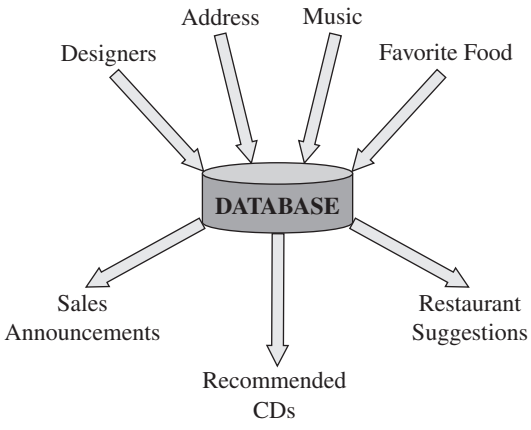
It soon became clear that a new kind of software was needed to help manage the data, as well as faster hardware to keep up with the increasing volume of data and data access demands. In terms of personnel, data management specialists would have to be developed, educated, and given the responsibility for managing the data as a corporate resource.

Out of this need was born a new kind of software, the database management system (DBMS), and a new category of personnel, with titles like database administrator and data management specialist. And, yes, hardware has progressively gotten faster and cheaper for the degree of performance that it provides. The integration of these advances adds up to much more than the simple sum of their parts; they add up to the database environment.

1.1.3 The Need for Data Management

It is practically impossible to buy anything, sell anything, or travel anywhere by air, rail, or sea without the fact being recorded in a database somewhere. With the recent rash of mergers of all kinds of organizations into larger entities, this data is becoming centralized. Residing in databases like the one shown in Figure 1-3, the data can be "mined" for useful information, allowing companies to find out not only who you are and where you live, but also what you like to eat, what you like to read, and who your favorite musicians and entertainers are. They know what your favorite sports teams are, and what sports you like to participate in yourself. They know where you shop and how often. They know when you are about to run out of something you buy regularly. They know when your kids are born, when they

Figure 1-3



Database and mining results.

are about to enter kindergarten, when they will graduate from high school, and when they are engaged to be married. They then use the information generated from the data stored in the database, along with sophisticated mining software, to target sales efforts with a higher probability of success.

All this data is stored in databases. The databases are growing larger, not only because more data is added to them on a daily basis, but because new kinds of data are being captured and stored, based on the activities and transactions in which you participate in the course of your daily living. The amount of data being stored in databases every day, based on people's actions and transactions, is already huge, but will get even larger in the coming years.

FOR EXAMPLE

Amazon.com

Amazon.com, the largest retailer on the Web, has perfected the technique of using databases to characterize its customers. By analyzing the kinds of products you have bought or expressed interest in, Amazon.com can present you with displays of similar products that you are likely to find interesting. This sales strategy requires not only massive, well-structured databases, but also sophisticated data-mining software that finds associations and relationships in customers' past behavior in order to prediction what customers are likely to do and want in the future.

SELF-CHECK

- Explain why businesses see data management as a critical requirement.
- List companies that collect data about their customers as part of their normal business activities.

1.2 Understanding Data Sources

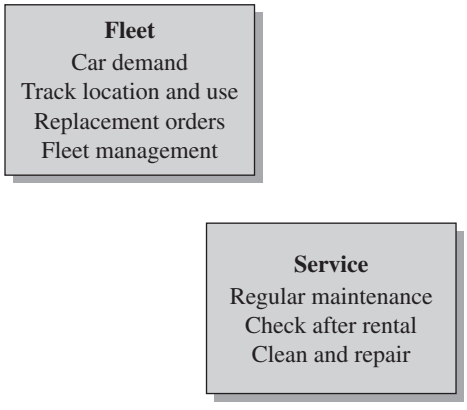
A single piece of data is a single fact about something that interests us. Think about the world around you, about your environment. In any environment there are things that are important to you, and there are facts about those things that are worth remembering. A thing can be an obvious object like an automobile or a piece of furniture. But the concept of an object is broad enough to include a person, an organization like a company, or an event that took place, like a particular meeting. A fact can be any characteristic of an object. In a university environment it may be the fact that student Gloria Thomas has completed ninety-six credits, or it may be the fact that Professor Howard Gold graduated from Ohio State University, or it may be the fact that English 349 is being held in Room 830 of Alumni Hall. In a commercial environment, it may be the fact that employee John Baker's employee number is 137; or it may be the fact that one of a company's suppliers, the Superior Products Company, is located in Chicago; or it may be the fact that the refrigerator with serial number 958304 was manufactured on November 5, 2004. Usually, we have many facts to describe something of interest to us in our environment.

1.2.1 Picking a Starting Point

Before you can begin organizing data into some useful form, available data must be collected. Data gathering can be tedious, but is critical. Initially, you aren't concerned about sorting the data or trying to identify what is or isn't important. Instead, you need to just get the data together in one central location. Once you have that, you can start figuring out what you can (and should) do with it.

A good place to start is with identifying the company's **core business**—its primary activity and reason for existence. Notice the numerous activities in a company. For example, look at the activities in a car rental business. Different departments are engaged in various activities, like the examples in Figure 1-4. What are the activities in the department that manages the fleet? This department studies the demand for different types of cars. It examines the usage and status at individual branches. It places orders to replace cars that need to be

Figure 1-4



Departmental activities.

retired soon. All these and many more activities within the fleet management department are aimed at supporting the core business of the car rental company, which is renting cars to customers. The servicing department's activities are focused on keeping each car in top condition. Each car must be thoroughly checked after each rental and maintained. Again, the activities of the servicing department, although different from those of fleet management, are also aimed at supporting the core business of the car rental company, namely, renting cars to customers.

Here is a sample of the core businesses of a few types of organizations:

- Retail grocery store: buy groceries from vendors and sell to retail customers.
- Stock brokerage: buy and sell stocks and bonds for individuals and institutions.
- Auction company: enable customers to sell and buy goods through auctions.
- Computer consulting: provide consulting services.
- Airlines: provide air transportation to customers.
- Car dealership: buy and sell cars.
- Department store: buy and sell consumer goods.
- University: provide higher learning to students.

After you identify what a business is and what it does, you can start to understand the data involved.

1.2.2 Identifying Primary Processes

The core business of an organization typically is to provide goods or services to its customers. The functions of each person and every department in the organization

are directed toward accomplishing the purpose of the core business. Many distinctive processes carried out by the departments support the core business. These primary processes fulfill the purpose of the core business.

At this point, data collection can still look like an overwhelming process. Even a simple business can easily have dozens, if not hundreds, of primary processes needed to accomplish its core business goals. Trying to address all of these processes at once is not only daunting, it's nearly impossible and, in the long run, a waste of time. It's better to start by focusing on one process, like airline ticketing and seat assignment.

What tasks must be completed to complete the process? To carry out the various processes, the company needs resources and assets such as buildings, equipment, materials, people, and money. But that is not all. The company also needs information to accomplish its processes. Information is a major asset, like other tangible and intangible assets of the company, used for performing the multitude of processes. Look for the information needed to complete the task, and you have a start on collecting the company's core information resources.

1.2.3 Specific Data Sources

Information comes in a wide variety of forms. In most cases, there is no shortage of available information from which to choose. However, these data sources are not always immediately obvious. You have to work with an organization to locate the information and dig it out from its many possible hiding places.

Finding Current Information

A good place to start is business documents. These include both hard copy and electronic documents. Some are common to most organizations, such as employee records, while others are more specific to the type of organization. Examples of business documents include items such as:

- Employee records: includes hiring and firing records, time sheets and other payroll information, vacation records, and related records.
- Customer records: includes customer lists, contact information, purchase history.
- "Hard" assets: long-term organizational assets such as land, buildings, and equipment.
- Inventory records: includes inventory of items for sale as well as items for internal consumption.
- Accounting records: information about the organization's resources with a focus on financial resources.

Some of these documents, such as accounting records, are likely already organized to some extent and in an electronic format. Though they often cannot be

used as is, they are typically easier to integrate with your database. Other documents may exist only as a hard-copy paper trail, leaving much of the organization (and data entry) up to you.

Finding Historic Information

The task in finding historic information is often not so much identifying the information you need as it is physically locating the information. The older the information, the greater the chance that it exists as paper documents only. These documents may be hidden away in old filing cabinets or in boxes somewhere in the back of a warehouse.

Sometimes you will find that there are gaps in the information, with documents lost, damaged, or destroyed. Some information might no longer apply to the core business activities. The organization may need to make a decision about how far back you need to dig through the historical record, as well as what part of that record is really important.

Be careful not to dismiss data as no longer useful just because it is old. Modern data mining software can find unsuspected relationships between old and new data, identify unexpected trends, and extract a treasure trove of business intelligence.

Using Human Sources

Employees at all levels of the organization are a vital information source during data collection. However, it's important to understand that the information provided by an employee is colored by the employee's environment. Employees are most concerned about the data they need to do their jobs, as in Figure 1-5.

Supervisors and executive staff usually focus on the big picture. They can give you an overall understanding of the organization and how things fit together. They can also provide insight into organization goals, which can directly impact

Figure 1-5

Executive	Big picture data Company overview
Managers	Day-to-day activities
Employees	Job requirements Task-specific information

Human data sources.

data requirements. Front-line managers are more concerned about day-to-day activities, so this is the data with which they are most familiar. A warehouse manager, for example, can provide information about inventory and about purchasing procedures.

Often, the richest source for data requirements and practical data are rank-and-file employees. You need to spend time talking with employees as part of any data collection process. Individual employees might not be able to describe exactly what information they need to do their jobs, but they can describe the jobs they do in detail. From this, you can get an understanding of the data they need. You also get insight into how they need to have the data made available, which can be just as important when designing and developing databases and data applications. You can usually get the most complete picture of the tasks an employee performs through direct observation. Watch what they do and how they do it. If you don't understand something or why it is being done a particular way, ask.

Understanding Data Flows

Part of the data collection process includes understanding how data flows through an organization. What are the data sources? Where or how is the data generated? The answers to these questions often help you locate the data in its purest form.

Data is similar to water, flowing from source to destination. However, with water you have streams running together into rivers, then into larger rivers, and finally into the ocean as its final destination. Data, as it travels, is likely to continue branching out to more and more locations. This is important to you, though, because the better you understand the data destinations (often referred to as **data consumers**), the better you understand both data organization and access requirements. It also helps you design and manage data access security.

FOR EXAMPLE

Hands-on Collection

You are gathering data for a retail organization. How do you find out the data requirements for supporting telephone sales? You do so by watching the sales personnel in action. Watch for both standard requirements and exceptions. How do sales personnel verify product information? How do they check stocking levels? What do they record about the products sold? What do they record about the customer? What is required to complete the transaction? These questions all help to provide direction to your data collection efforts.

SELF-CHECK

- Explain how primary processes can be used to identify data collection requirements.
- Explain why a company's organization chart can be important in determining the employees to interview and what questions to ask during data collection.
- Take a business model with which you are familiar, such as retail grocer, department store, or a school. List the primary processes needed to accomplish core business goals and the types of data needed by each.

1.3 Potential Data Concerns

Every corporate resource has to be carefully managed in order for a company to keep track of it, protect it, and distribute it to those people and purposes in the company that need it. Furthermore, public companies have a responsibility to their shareholders to competently manage the company's assets. The chief financial officer, and a staff of accountants and financial professionals, is responsible for the company's financial resources, with outside accounting firms providing independent audits. Typically, vice presidents of personnel and their staffs are responsible for the administrative functions that are necessary to manage employee affairs; production managers at various levels are responsible for parts inventories; and so on. Data is no exception: it is a resource that must be carefully managed.

A database is no better than the data it contains. The data, in turn, is no better than its source. Business documents are not always accurate or up-to-date. Employees, from rank-and-file through executive staff, can be misinformed or might not understand your question. Sometimes, they might even lie. It isn't always possible to verify all of the data you collect. Still, you need to be aware of the potential for inaccuracies and be ready to correct your data as necessary. One advantage of data management systems is that some types of data inconsistencies become immediately obvious.

Data accuracy is just one of your potential concerns when working with data. You also need to consider data security. Data, as already discussed, is a critical resource. What happens if a competitor gets access to that resource, or if someone accidentally (or maliciously) deletes or modifies the data? Even if your data is accurate and secure, you also have to make it available to those that need it, while maintaining adequate security. This means having the data properly organized and accessible to those who legitimately require access.

1.3.1 Managing Data Accuracy

Data accuracy (ensuring that data is accurate and correct) begins with verifying your source and continues with careful data entry and data manipulation. Most data management systems provide tools to help ensure data accuracy through various means, such as establishing relationships between data values and setting appropriate limits on data values.

During data collection, keep in mind that even though human data sources are an important resource, they are not always accurate. The information provided by employees is filtered by the context of the job they perform and the temptation to put oneself in the best possible light. People often completely ignore some functions they perform simply because they are too familiar with their jobs. They might overlook critical steps or leave out details because they don't seem important. Occasionally, employees at different levels might feel a need to purposely hide information for various reasons, such as the belief that doing so will increase the employee's job security (if the employee is the only one that knows key information, then it's more difficult for the company to fire him or her).

You need to validate your data sources, as far as is possible. You must sometimes also apply a lifetime to the data, a description of how long the data should be considered as accurate. This can be a formal designation or more informal, based a sliding value judgment. Inaccuracies can also arise through poorly designed algorithms (formulas) used to manipulate data.

1.3.2 Managing Data Security

The term data security is somewhat flexible, depending on the context in which it is used. Say "data security" to one person, and the first things that come to mind are protecting the data against unauthorized access and modification. To someone else, it means making sure that the data is protected against corruption or loss. Both are critical concerns. In general, **data security** refers to processes and procedures implemented with the goal of keeping data safe. Basic data security categories include access security and physical security.

The tools available for ensuring data security are somewhat application- and platform-specific. This is important because data is a tempting target. Many of the most destructive viruses, Trojan horses, and other malicious software floating around the Internet right now directly target database applications.

Considering Access Security

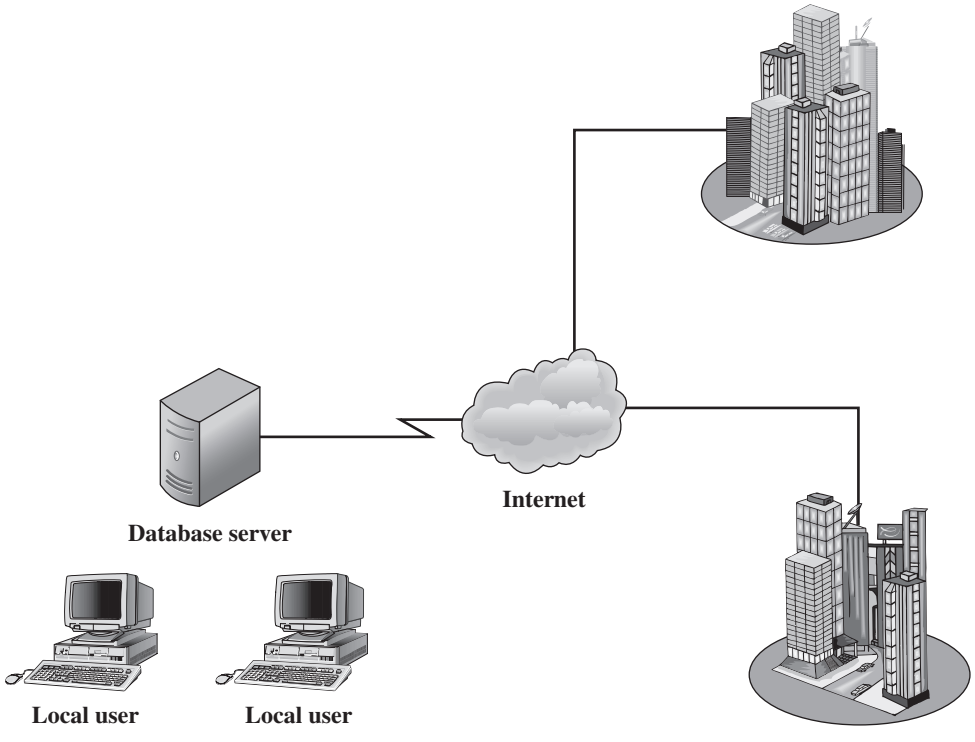
Most database systems control **access security** (controlling user and application access to data) through support of one or more **authentication methods**, which are ways of verifying a user's identity before allowing access to the data. Different levels of access permissions are also provided. For example, you can give a

user permission to view data, but not change it. Accessibility is also an issue when discussing security. You want to protect the data, but you also need to make it available to the appropriate users.

A vital part of this effort is reducing the **surface area** (a term used to refer to how exposed a database is to access and manipulation) that is vulnerable to attack. The wider your **consumer audience** (or data consumers—the people or applications who need access to data and who use data in some manner), the more difficult this task becomes. Consider a local database that contains all of your company's data. If your employees are the only people that need access to the data, protecting the data is a relatively easy task. The issue of who needs access, as well as the level of access required, is typically well defined. In many cases, you can implement many of the security requirements as an extension to existing network security.

As your potential consumer audience increases, so do your potential security concerns. What if people outside of your company need access to the data, possibly accessing your network resources remotely through the Internet, as shown in Figure 1-6?

Figure 1-6



Remote consumer audience (user population).

Now you have less control over the user population. Though not shown in Figure 1-6, additional tools such as network firewalls often come into play. Rather than providing direct access to the data like you have in this example, you might instead provide indirect access, such as through a Web-based application. You will find that there are no one-size-fits-all solutions. Instead, it's likely that you will need to design a custom-access solution matched to the specific access requirements, and be ready to modify the solution and access as needs evolve.

Considering Physical Security

Protecting the data against physical corruption or loss is the role of **physical security.** This usually comes down to three basic tasks:

- Protecting the server
- Ensuring the storage media
- Duplicating the data

Protecting the server is usually a matter of providing physical security, locking the server away from prying eyes and busy fingers. A dedicated, secure server room is the rule in most organizations, with access to server hardware strictly limited.

Ensuring the storage media means making sure that it is reliable and provides a safe location for your data. One of the most common ways of doing this is through fault-tolerant disk systems. A **fault-tolerant disk system,** usually based on Redundant Array of Independent Disk (RAID) technologies, is designed to protect the data in the event of a hard disk failure. Usually, when a hard disk fails, all data on the disk is lost. RAID systems use different configurations based on multiple hard disks that keep the data available and protected should any one hard disk fail. When this happens, it's usually an easy matter to replace the failing disk and reestablish the data protection.

Duplicating data can be accomplished through making a literal duplicate copy, or **mirror image,** of the data. You could keep a copy of the database on one or more additional servers. However, these data image solutions can be expensive to implement and difficult to maintain. It's more likely that you implement the duplication of data through data backups, data copied to a different storage destination and available should data loss occur. You can back up the data to another computer, such as a network file server, or to removable media such as magnetic tape or DVD disk. If something does happen and the database is lost or corrupted, you restore the data from the backup, recreating the database and its data. Of course, any data changes made since the last time you ran a data backup will be lost and must be reposted to the database manually.

1.3.4 Managing Data Organization

For the huge amount of data being captured to be useful, it must be organized. Organization forces a structure on the data, helps to reduce the amount of physical space the data requires, and can help to minimize the time required to locate and retrieve the data that you want. In fact, your only hope for making sense out of the flood of data is to organize it in a way that allows fast, efficient retrieval of the information you want, regardless of how large the dataset is. The longer you wait to organize your data, the harder it will be to do. Whether you are just starting out and as yet have not collected any data, or whether you are in an established organization that has been collecting data for years, there will never be a better time than right now to decide the best way to organize your data so that you can quickly receive answers to the questions you will want to ask both now and in the years to come. After you decide what kinds of information you are likely to want to extract from the data, you can design a database that will make it easy to do so.

Data standards also play a critical role. For the data to be useful, it needs to be organized and entered in the same way for each data object. You need to collect the same data about each instance of an object, and store it the same way each time. For example, when collecting employee data, you would use the same format for storing each employee's name, employee ID, hiring data, and other information about that employee. You need to refer to the data using the same terminology to avoid confusing data consumers and application programmers.

Because of its importance, organization will be your next major task after data collection. This means identifying what is important and deciding how to represent it in a database. You must identify the key items that you want to track, typically referred to as entities, and the information that you want to keep about them, typically referred to as attributes. However, this is just the start of the process. You will need to consider how different entities influence and relate to one another, account for data that depends directly on other data, and so on. Only then can you start seriously designing a database application.

1.3.5 Managing Data Access

We've already mentioned **data access,** the process of allowing users or applications to retrieve data from a database. After all, a database isn't of any use to anyone if they can't retrieve the information that they need from it. Your data security design directly impacts user access. When you organize your data, one of the goals is to optimize performance when accessing and retrieving data.

As you prepare to build your database application, you'll spend much of your time deciding on the database objects you need to include in the database, both for data storage and to support data retrieval. Most solutions also mean programming custom executable objects to support data entry, manipulation, and retrieval.

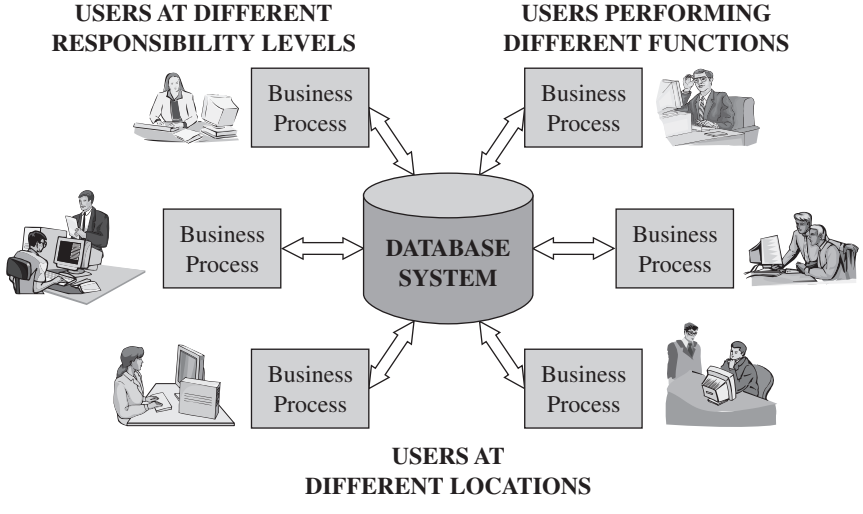
The term commonly used to refer to data access is a **data query.** A query, though it often implies simple data retrieval, can also be used to insert new data or manipulate existing data. Most database systems use a standard set of commands known as the Structured Query Language, or SQL, to query database data. That is why you often hear modern databases referred to as SQL databases.

When you set up your database, you must determine who needs access to the data and what level of access they need. Not all users need access to the same data. Access requirements are related to users' responsibilities and the business functions (primary practices) they perform (Figure 1-7). Think back to the processes relating to airline ticketing and seat assignment. Ticket agents need access to a great deal of information such as flights, available seats, and ticket prices. There is even more information about the airline, though, to which the ticket agent should not have access, such as confidential financial records.

In addition to determining who needs access, another design decision you'll need to make fairly early is exactly how to give users access to the database. Typically, a very limited number of users have direct access to the database. This is usually restricted to database administrators and, possibly, database programmers. Nearly everyone else accesses the database indirectly through another application. The specific application type depends on the access requirements.

Users responsible for day-to-day activities typically access a database through one or more applications designed to perform specific tasks. For example, ticket

Figure 1-7



Information access requirements by responsibility and role.

FOR EXAMPLE

The Galileo Spacecraft and Its Data

The Galileo spacecraft has been studying the Jupiter system for several years. It has sent back huge amounts of data. By studying a small fraction of that data, scientists have inferred the existence of a global ocean under the ice covering the surface of Jupiter's moon Europa and the probable existence of a similar ocean under the surface of another of Jupiter's moons, Callisto. However, the large majority of Galileo's data remains unstudied, because in its unorganized state, it is very difficult to extract useful information from it, unless you already know what you are looking for. Organizing the data into a database would be of tremendous benefit, but NASA has no funding for such a massive effort.

agents for an airline would have an application designed to sell tickets and assign seats. A user in a department store working a point-of-sale terminal would have an application designed to post to sales and inventory records. People working in the accounting department would have an application designed to work with financial accounting records. Application requirements relate back to the core business and primary processes.

Other access requirements could add another layer of isolation between the data consumer and the database. Rather than communicating directly with the database, an application might send and receive data through a Web service, which, in turn, is the application that communicates with the database.

SELF-CHECK

- Lists key concerns relating to data management.
- Explain how your network environment can influence your access and security design.
- Explain the need for different levels of data-access security based on responsibilities and roles.
- Identify tools you might need to implement physical data security.
- Compare and contrast direct database access, indirect access by local clients, and indirect access by remote clients.

SUMMARY

In this chapter, you learned fundamental concepts relating to data, its role as a business resource, and requirements for data management. You compared data requirements from both historic and modern perspectives, including a look at a specific business process and related data requirements. You investigated the need for data collection and how to identify data sources based on core business activities and primary processes. You were also introduced to some of the fundamentals of data management and potential concerns related to data management.

KEY TERMS

Access security
Authentication method
Consumer audience
Core business
Data
Data access
Data accuracy
Data consumer
Data query
Data security
Database
Fault-tolerant disk system
Information
Mirror image
Physical security
Surface area

ASSESS YOUR UNDERSTANDING

Go to www.wiley.com/college/gillenson to evaluate your knowledge of data and data management.
Measure your learning by comparing pre-test and post-test results.

Summary Questions

1. The terms data and information mean the same thing. True or false?
2. What is a database management system?
   - **(a)** an organization responsible for database maintenance and administration
   - **(b)** specialized software designed to implement, support, and maintain databases
   - **(c)** a manual method of record-keeping phased out with the introduction of computers
   - **(d)** an end-user application designed to give users access to data
3. Which statement best describes the role of data in business?
   - **(a)** Data collection and organization is a low-priority effort with few benefits for a business.
   - **(b)** Data is important to business in the information technology sector only.
   - **(c)** All businesses freely share data about their customers in a cooperative environment.
   - **(d)** Data can provide a competitive advantage in hotly contested industries.
4. Which of the following statements best describe the use of business documents during initial data collection?
   - **(a)** You should collect all business documents except those relating to personnel records.
   - **(b)** Limit collection efforts to documents in electronic format only as hard-copy documents seldom contain useful information.
   - **(c)** You should collect all available business documents, including both electronic and hard-copy documents.
   - **(d)** It is safe to assume that all business documents you locate are complete, accurate, and applicable to the current business environment.
5. When interviewing employees to collect data, you should limit yourself to executive staff, supervisors, and managers only. True or False?
6. A good place for identifying the starting point for data collection is an organization's core business. True or False?

7. How do primary processes relate to data collection activities?
   (a) Primary processes give you a way of validating data accuracy.
   (b) The tasks required to complete primary processes identify critical business data sources.
   (c) Primary processes DO NOT relate to data collection activities.
8. Executive staff personnel are a good source for what kind of information?
   (a) an overall understanding of the organization and its goals
   (b) critical details of day-to-day activities at all levels
   (c) details of the tasks required to perform primary processes
   (d) identification of data paths within the organization
9. How do primary processes relate to an organization's core business?
   (a) The primary processes and core business are one in the same.
   (b) The primary processes fulfill the core business.
   (c) The primary processes are those business activities occurring outside of the core business.
   (d) The primary processes provide a clear and concise definition of the core business.
10. Which of the following is an advantage of records in an electronic format versus those in hard copy only?
   (a) The records relate to current business activities.
   (b) The records can typically be used as is, without additional modification.
   (c) The records are typically already organized for you.
11. Which of the following is typically NOT a goal in data management?
   (a) keeping data physically secure
   (b) making all data available to all employees
   (c) ensuring that the data is timely and accurate
   (d) ensuring that the data is organized for easy retrieval
12. Raw data is just as useful and valuable as organized data. True or False?
13. For what purpose are authentication methods implemented?
   (a) to verify users before allowing access to data
   (b) to set access limits through defined permissions
   (c) to provide physical security for critical data
   (d) to maintain data organization
14. For what reason would you implement RAID technologies as part of your design?
   (a) to provide user authentication
   (b) to limit user access permissions

(c) to provide fault-tolerant data storage
(d) to organize data for easy retrieval

15. Which of the following is NOT a reason for organizing data in a database?
(a) to provide fault tolerant storage
(b) to reduce the space required for storage
(c) to optimize data retrieval times
(d) to force a structure on the data

16. Database backups are an example of which of the following?
(a) data organization
(b) access security
(c) physical security
(d) user access methods

Applying This Chapter

1. You are trying to justify a computerized data management system to your supervisor. You work in a service industry that deals directly with customers. How could collecting information about your customers provide a competitive edge? What kind of software would you need to detect trends and make projections about customer activities?

2. You have been asked to design a database application for a company in the retail industry. The company sells through physical stores, through telephone sales, and through e-commerce over the Internet. How can you determine the best sources of information for each of the sales efforts? How can you determine what information you need to collect from each of the sales efforts?

3. Your company has implemented a database management system to track all business-related data from the company. Access to the data should be limited to employees only. How can you ensure this? What other access protections should you place on the data? What guidelines should you use in determining levels of user access? What methods can you use to protect the data against loss or corruption?

YOU TRY IT

Data Collection and Organization

You are heading up the data collection efforts for your company's new database project. You are responsible for determining what information you need to collect and how it should be collected.

1. Discuss how you can identify the primary processes within a company.
2. What might the company have available that would let you associate employees with the jobs they perform?
3. Why is it important to collect all available information during the initial collection process?

Data Management

You are responsible for identifying data management concerns within your organization. Your company includes employees that telecommute and must access the local network via the Internet. Both local employees and telecommuters will need access to data in the database.

1. Explain why it is important to both implement access security and assign access permissions to data. What guidelines should you use for assigning permissions?
2. Discuss possible options for providing employees with access to data. Compare available options for local employees and telecommuters.