

WARNING CONCERNING COPYRIGHT RESTRICTIONS

Under certain conditions specified in the United States Copyright law (Title 17, United States Code) libraries and archives may be authorized to furnish a photocopy or other reproduction. One of these conditions is that the photocopy or reproduction may not be 'used for any purpose other than private study, scholarship, or research' and that only one copy be provided for scholarly purposes, unless copyright fees are paid. Further reproduction of paper copies made from this computer system may be in violation of copyright laws and is prohibited.

MANNS.P05A

32

PLEASE TYPE, OR PRINT LEGIBLY, IN BLACK INK - THIS PAGE MAY BE SCANNED

RAMSEY LIBRARY RESERVE MATERIALS FORM

PHOTOCOPIES

THIS ITEM TO BE PLACED ON:

- TRADITIONAL RESERVE (ALLOW 48 HOURS TO PROCESS)
- ELECTRONIC RESERVE (ALLOW 1 WEEK TO PROCESS)

TITLE AS YOU WANT IT LISTED: DATABASES AND DATA Warehouses

DATE 1/13/05 FALL
INSTRUCTOR M. L. MANNS SPRING
COURSE NUMBER MGMT 386 SUMMER
COURSE NAME Systems & Info Mgmt
INSTITUTION UNCA

Please Note: This form must accompany each reserve article

In compliance with copyright law, the 1st page of any reserve article *must* contain the bibliographic information requested below, in full. Fill out a separate sheet for each item you wish to be placed on reserve.

() Title of Periodical _____
Title of Article _____
Author _____ Year _____
Volume _____ Issue _____ Pages _____

() Title of Book MANAGEMENT Information Systems
Article Title or Chapter DATABASES & DATA Warehouses
Author Haag, Cummings, McCubbrey
Edition/Date of Publication 5th Publisher McGraw-Hill
Pages 125-155

CHAPTER THREE

Databases and Data Warehouses Building Business Intelligence

OPENING CASE STUDY: CHRYSLER SPINS A COMPETITIVE ADVANTAGE WITH SUPPLY CHAIN MANAGEMENT SOFTWARE

According to John Kay, DaimlerChrysler's manager of electronic commerce, "Being able to get critical product information to our external suppliers as soon as it becomes available is a definite competitive advantage for all concerned." John is talking about Chrysler's Supply Partner Information Network (SPIN), a Web-based supply chain management system that increased productivity by 20 percent and reduced operating costs in the first year of its implementation.

The two technology tools critical to the success of SPIN, which already boasts over 3,500 suppliers accessing and using it, are the Internet and databases. The Internet, of course, allows 24 × 7 access to SPIN by Chrysler's suppliers for providing parts, bidding on contracts, and submitting purchase invoices. The Internet also allows Chrysler to send to its suppliers in real time vitally important information updates on procurement requirements, strategy applications, and the like.

What's internal to SPIN is a powerful set of databases that track, organize, and maintain all of Chrysler's inventory, material requirements planning, work-in-progress, and supplier information. A database is a technology tool that enables you to organize vast amounts of information in the most logical way that suits your business needs. For SPIN to be successful, databases are an absolute necessity. As Jeremy Hamilton-Wright, the team leader in Chrysler's IS department, describes it, "SPIN supports everything from developing products to delivering parts and sending payments. SPIN works for all of Chrysler's different types of suppliers: production suppliers,

parts suppliers, and the suppliers that package parts."

All businesses today are using databases to organize and manage their information. Why? Because databases allow you to create logical relationships within your information. At Chrysler, for example, recall notices for a defective part or assembly are logically tracked back through the production process, including human operators and equipment, through database information. If Chrysler determines that the error didn't occur in the production process, it tracks information further back through its databases to its inventory and suppliers.

Many organizations also use databases in support of their customer relationship management (CRM) strategies, another source of competitive advantage. Using CRM-enabled databases, organizations today can track customer purchases, purchases by credit card or cash, purchases by time of the day and day of the week, purchases by store location, and an array of customer demographics. All that information can then be logically organized in such a way that you could easily find the answer to the following question: "Which of our customers who have purchased over \$1,500 in the last six months at a store within a five-mile radius of their homes have a household income above \$75,000 and no children?"

Answers to those types of questions are what we call *business intelligence*. Business intelligence helps you make sense of your information by allowing you to view it from different perspectives and ask insightful and thought-provoking questions. To support the creation of business intelligence, many organizations are taking information from their databases and creating data warehouses, another technology tool that supports the logical organization of information.¹

Introduction

As we've discussed in the first two chapters, you and your organization need more than just data and information. You need **business intelligence (BI)**—knowledge about your customers, your competitors, your business partners, your competitive environment, and your own internal operations—that gives you the ability to make effective, important, and often strategic business decisions. It enables your organization to extract the true meaning of information so that you can take creative and powerful steps to ensure a competitive advantage. Many such actions by your organization support some or all the initiatives we discussed in Chapter 2—customer relationship management, supply chain management, and collaboration, to name just a few.

Of course, to create business intelligence you need both data and *information* (we'll commonly refer to both as *information* in this chapter). So, business intelligence doesn't just magically appear. You must first gather and organize all your information. Then, you have to have the right IT tools to define and analyze various relationships within the information. In short, knowledge workers such as you use IT tools to create business intelligence from information. The technology, by itself, simply won't do it for you. However, technology such as databases, database management systems, data warehouses, and data-mining tools can definitely help you build and use business intelligence.

As you begin working with these IT tools (which we'll discuss in great detail throughout this chapter), you'll be performing the two types of information processing we alluded to in Chapter 1: online transaction processing and online analytical processing. **Online transaction processing (OLTP)** is the gathering of input information, processing that information, and updating existing information to reflect the gathered and processed information. Databases and DBMSs are the technology tools that directly support OLTP. Databases that support OLTP are most often referred to as **operational databases**. Inside these operational databases is valuable information that forms the basis for business intelligence.

As you can see in Figure 3.1 (on the facing page), you can also query operational databases to gather basic forms of business intelligence, such as how many products individually sold over \$10,000 last month and how much money was spent last month on radio advertising. While the results of these queries may be helpful, you really need to combine product and advertising information (with several other types of information including customer demographics) to perform online analytical processing.

Online analytical processing (OLAP) is the manipulation of information to support decision making. At Australian P&C Direct, OLAP within a data warehouse is a must. P&C has created a data warehouse that supports its customer relationship management activities, cross-selling strategies, and marketing campaigns. By creating a data warehouse with customer information (including census data and lifestyle codes), its wide array of insurance and financial products, and its marketing campaign information, P&C agents can view all the products a given customer has purchased and more accurately determine cross-selling opportunities and what marketing campaigns a given customer is likely to respond to.²

A data warehouse is, in fact, a special form of a database that contains information gathered from operational databases for the purpose of supporting decision-making tasks. When you build a data warehouse and use data-mining tools to manipulate the data warehouse's information, your single goal is to create *business intelligence*. So, data warehouses support only OLAP; they do not at all support OLTP. As you can see in Figure 3.1, you can perform more in-depth queries to gather business intelligence from a

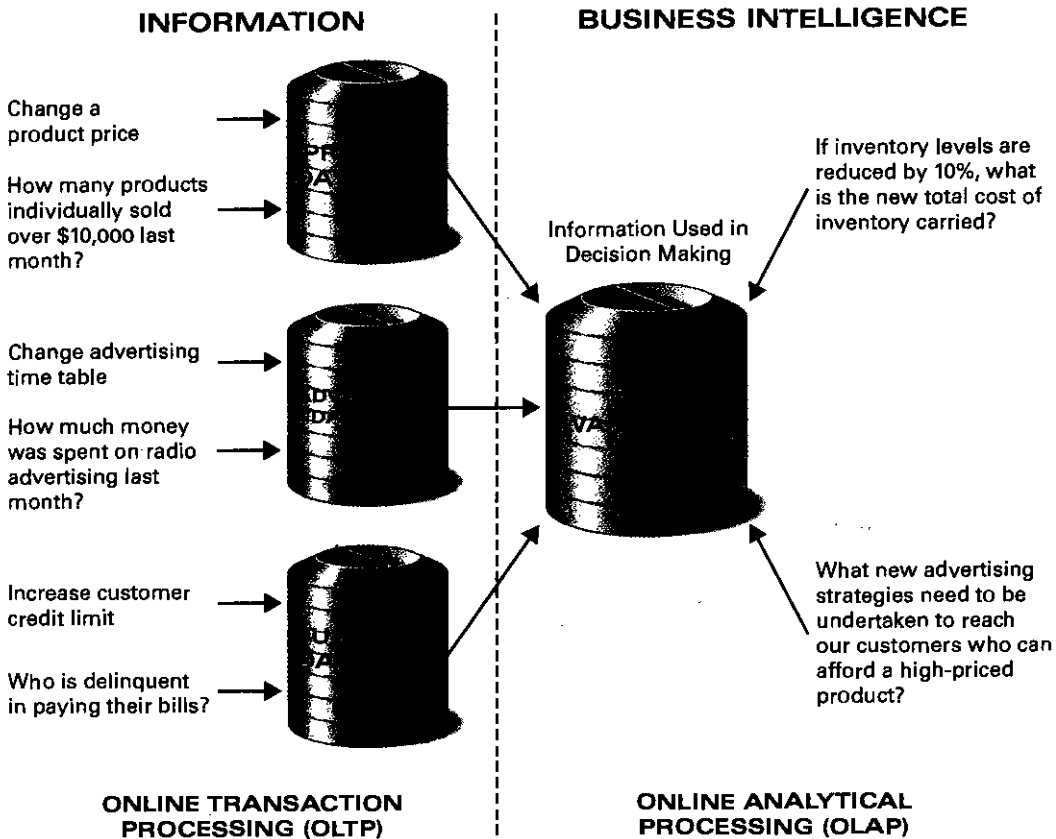


Figure 3.1

Building Business Intelligence

data warehouse than you can with a single database. For example, "What new advertising strategies need to be undertaken to reach our customers who can afford a high-priced product?" is a query that would require information from multiple databases. Data warehouses better support creating that type of business intelligence than do databases.

As this chapter unfolds, we'll look specifically at (1) databases and database management systems and (2) data warehouses and data-mining tools. Databases today are the foundation for organizing and managing information, and database management systems provide the tools you use to work with a database. Data warehouses are relatively new technologies that help you organize and manage business intelligence, and data-mining tools help you extract that vitally important business intelligence. Data warehouses and data-mining tools are a subset of the business intelligence software that we discussed in Chapter 2.

As we first look at databases and database management systems in this chapter, we'll be exploring their use by Solomon Enterprises in support of customer relationship management and order processing. Solomon Enterprises specializes in providing concrete to commercial builders and individual homeowners in the greater Chicago area. Solomon tracks detailed information on its concrete types, customers, raw materials, raw materials' suppliers, trucks, and employees. It uses a database to organize and manage all this information. As we discuss Solomon Enterprises and its use of a database, we'll focus mostly on CRM and ordering processing. In *Extended Learning Module C*, which follows this chapter, we'll look at how to design the supply chain management side of Solomon's database.

The Relational Database Model

For organizing and storing basic and transaction-oriented information (that is eventually used to create business intelligence), businesses today use databases. There are actually four primary models for creating a database. The object-oriented database model is the newest and holds great promise; we'll talk more about the entire object-oriented genre in Chapter 7 and in *Extended Learning Module G*. Right now, let's focus on the most popular database model: the relational database model.

As a generic definition, we would say that any **database** is a collection of information that you organize and access according to the logical structure of that information. In reference to a **relational database**, we say that it uses a series of logically related two-dimensional tables or files to store information in the form of a database. The term **relation** often describes each two-dimensional table or file in the relational model (hence its name *relational* database model). A relational database is actually composed of two distinct parts: (1) the information itself, stored in a series of two-dimensional tables, files, or relations (people use these three terms interchangeably) and (2) the logical structure of that information. Let's look at a portion of Solomon's database to further explore the characteristics of the relational database model.

COLLECTIONS OF INFORMATION

In Figure 3.2 (on the facing page), we've created a view of a portion of Solomon's database. Notice that it contains five files: *Order*, *Customer*, *Concrete Type*, *Employee*, and *Truck*. (It will contain many more as we develop it completely in *Extended Learning Module C*.) These files are all related for numerous reasons—customers make orders, employees drive trucks, an order has a concrete type, and so on. And you need all these files to manage your customer relationships and process orders.

Within each file, you can see specific pieces of information (or *attributes*). For example, the *Order* file contains *Order Number*, *Order Date*, *Customer Number*, *Delivery Address*, *Concrete Type*, *Amount* (this is given in cubic yards), *Truck Number*, and *Driver ID*. In the *Customer* file, you can see specific information including *Customer Number*, *Customer Name*, *Customer Phone*, and *Customer Primary Contact*. These are all important pieces of information that Solomon's database should contain. Moreover, Solomon needs all this information (and probably much more) to effectively process orders and manage customer relationships.

CREATED WITH LOGICAL STRUCTURES

Using the relational database model, you organize and access information according to its logical structure, not its physical position. So, you don't really care in which row of the *Employee* file Allison Smithson appears. You really need to know only that Allison's *Employee ID* is 984568756 or, for that matter, that her name is Allison Smithson. In the relational database model, a **data dictionary** contains the logical structure for the information in a database. When you create a database, you first create its data dictionary. The data dictionary contains important information (or logical properties) about your information. For example, the data dictionary for *Customer Phone* in the *Customer* file would require 10 digits. The data dictionary for *Date of Hire* in the *Employee* file would require a month, day, and year, as well.

This is quite different from other ways of organizing information. For example, if you want to access information in a certain cell in most spreadsheet applications, you must know its physical location—row number and column character. With a relational

ORDER FILE

Order Number	Order Date	Customer Number	Delivery Address	Concrete Type	Amount	Truck Number	Driver ID
100000	9/1/2004		55 Smith Lane		8		123456789
100001	9/1/2004	3456	2122 E. Biscayne	1	3	222	785934444
100002	9/2/2004		55 Smith Lane		6		435296657
100003	9/3/2004	4567	1333 Burr Ridge	2	4	333	435296657
100004	9/4/2004		1333 Burr Ridge		8		785934444
100005	9/4/2004	5678	1222 Westminster	1	4	222	785934444
100006	9/5/2004		222 East Hampton		4		123456789
100007	9/5/2004	2345	9 W. Palm Beach	2	5	333	785934444
100008	9/5/2004		4532 Lane Circle		8		785934444
100009	9/7/2004	1234	987 Furlong	3	8	111	123456789
100010	9/9/2004		4532 Lance Circle		7		435296657
100011	9/9/2004	4567	3500 Tomahawk	5	6	222	785934444

CUSTOMER FILE

Customer Number	Customer Name	Customer Phone	Customer Primary Contact
	Smelding Homes		Bill Johnson
2345	Home Builders Superior	3334444444	Marcus Connolly
	Mark Akey		Mark Akey
4567	Triple A Homes	3336666666	Janielle Smith
	Sheryl Williamson		Sheryl Williamson
6789	Home Makers	3338888888	John Yu

CONCRETE TYPE FILE

Concrete Type	Type Description
	Home foundation and walkways
2	Commercial foundation and infrastructure
	Premier speckled (concrete with pea-size smooth gravel aggregate)
4	Premier marble (concrete with crushed marble aggregate)
	Premier shell (concrete with shell aggregate)

EMPLOYEE FILE

Employee ID	Employee Last Name	Employee First Name	Date of Hire
	Johnson		2/1/1985
435296657	Evaraz	Antonio	3/3/1992
	Robertson		6/1/1999
984568756	Smithson	Allison	4/1/1997

TRUCK FILE

Truck Number	Truck Type	Date of Purchase
	Ford	
222	Ford	12/24/2001
333	Chevy	

Figure 3.2

A Portion of Solomon Enterprises' Database for Customer Relationship Management and Ordering Processing

database, however, you need only know the field name of the column of information (for example, *Amount*) and its logical row, not its physical row. As a result, in Solomon's database example, you could easily change the amount for an order, without having to know where that information is physically stored (by row or column).

And with spreadsheet software, you can immediately begin typing in information, creating column headings, and providing formatting. You can't do that with a database. Using a database, you must clearly define the characteristics of each field by creating a data dictionary. So, you must carefully plan the design of your database before you can start adding information.

WITH LOGICAL TIES WITHIN THE INFORMATION

In a relational database, you must create ties or relationships in the information that show how the files relate to each other. Before you can create these relationships among files, you must first specify the primary key of each file. A **primary key** is a field (or group of fields in some cases) that uniquely describes each record. In Solomon's database, *Order Number* is the primary key for the *Order* file and *Customer Number* is the primary key for the *Customer* file. That is to say, every order in the *Order* file must have a unique *Order Number* and every customer in the *Customer* file must have a unique *Customer Number*.

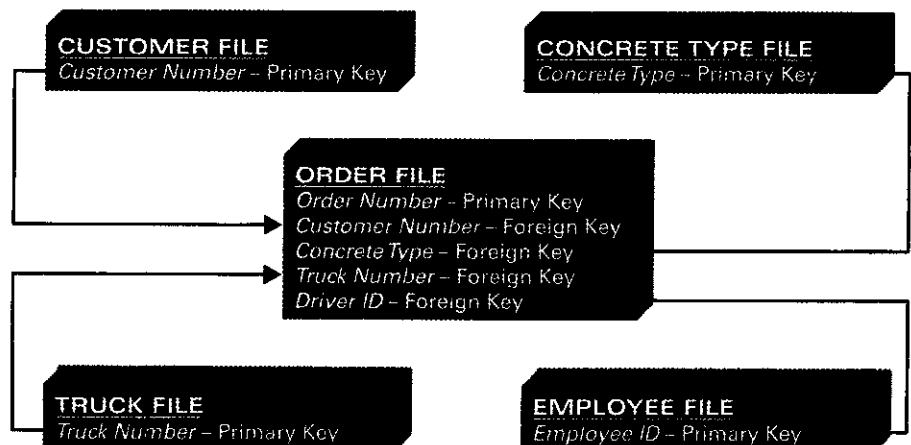
When you define that a specific field in a file is the primary key, you're also stating as well that the field cannot be blank. That is, you cannot enter the information for a new employee in the *Employee* file and leave the *Employee ID* field blank. If that were possible, you could potentially have two employees with identical primary keys (blank), which is not possible in a database environment.

Again, this is quite different from working with spreadsheets. Using a spreadsheet, it would be almost impossible to ensure that each field in a given column is unique. This reinforces the notion that, while spreadsheets work with information according to physical location, databases work with information logically.

If you look back at Figure 3.2, you can see that *Customer Number* appears in both the *Customer* and *Order* files. This creates a logical relationship between the two files and is an example of a foreign key. A **foreign key** is a primary key of one file that appears in another file. Now look at Figure 3.3. In it, we've provided the logical relationships among all five files. Notice, for example, that *Truck Number* is the primary key for the *Truck* file. It also appears in the *Order* file. This enables Solomon to track which trucks were used to deliver the various orders. So, *Truck Number* is the primary key in the *Truck* file and is also a foreign key that appears in the *Order* file. There are other examples of foreign keys as well in Figure 3.3.

Figure 3.3

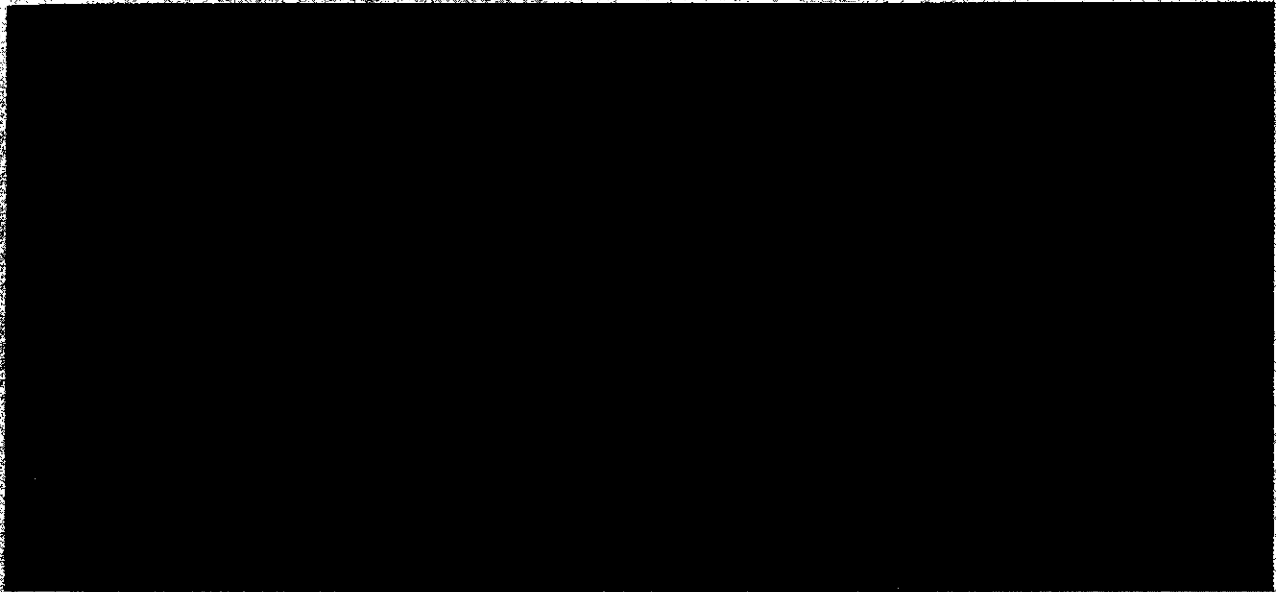
Creating Logical Ties with Primary and Foreign Keys



PRIMARY KEYS, FOREIGN KEYS, AND INTEGRITY CONSTRAINTS

Let's consider the information that your school tracks for a class. In this instance, a class is a scheduled course. For example, your school may have FINA 2100—Introduction to International Financial Markets as a course. If the school offers it in the fall, then it becomes a class. Below, we've provided many pieces of information that your school probably tracks about the class. First, which is the primary key (place an X in the second column)? Second, for each piece of information, identify if it's a

foreign key (a primary key of another file). If it is, write down the filename in the third column. Finally, in the fourth column for each piece of information, write down any integrity constraints you can think of. For example, can it be blank or must it contain something? Can it be duplicated across multiple records (classes)? If it's a number, does it have a specific range in which it must fall? There are many others.



Foreign keys are essential in the relational database model. Without them, you have no way of creating logical ties among the various files. As you might guess, we use these relationships extensively to create business intelligence because they enable us to track the logical relationships within many types of information.

WITH BUILT-IN INTEGRITY CONSTRAINTS

By defining the logical structure of information in a relational database, you're also developing *integrity constraints*—rules that help ensure the quality of the information. For example, by stating that *Customer Number* is the primary key of the *Customer* file and a foreign key in the *Order* file, you're saying (1) that no two customers can have the same *Customer Number* and (2) that a *Customer Number* that is entered into the *Order* file must have a matching *Customer Number* in the *Customer* file. So, as Solomon creates a new order and enters a *Customer Number* in the *Order* file, the database management system must find a corresponding and identical *Customer Number* in the *Customer* file. This makes perfect sense. You cannot create an order for a customer who does not exist.

Consumer Reports magazine has rated the Ritz-Carlton first among luxury hotels.³ Why? It's simple: Ritz-Carlton has created a powerful guest preference database to provide customized, personal, and high-level service to guests of any of its hotels. For example, if you leave a message at a Ritz-Carlton front desk that you want the bed turned down at 9 P.M., prefer no chocolate mints on your pillow, and want to participate in the 7 A.M. aerobics class, that information is passed along to the floor maid (and others) and is also stored in the guest preference database. By assigning to you a unique customer ID that creates logical ties to your various preferences, the Ritz-Carlton transfers your information to all of its other hotels. The next time you stay in a Ritz-Carlton hotel, in Palm Beach for example, your information is already there, and the hotel staff immediately knows of your preferences.

For the management at Ritz-Carlton, achieving customer loyalty starts first with knowing each customer individually (the concept of customer relationship management). That includes your exercise habits, what you most commonly consume from the snack bar in your room, how many towels you use daily, and whether you like a chocolate on your pillow. To store and organize all this information, Ritz-Carlton uses a relational database, and employees use it to meet your needs (or whims).

Database Management System Tools

When working with word processing software, you create and edit a document. When working with spreadsheet software, you create and edit a workbook. The same is true in a database environment. A database is equivalent to a document or a workbook because they all contain information. And while word processing and spreadsheet are the software tools you use to work with documents and workbooks, you use database management system software to work with databases. A *database management system (DBMS)* helps you specify the logical organization for a database and access and use the information within a database. A DBMS contains five important software components (see Figure 3.4):

1. DBMS engine
2. Data definition subsystem
3. Data manipulation subsystem
4. Application generation subsystem
5. Data administration subsystem

The DBMS engine is perhaps the most important, yet seldom recognized, component of a DBMS. The *DBMS engine* accepts logical requests from the various other DBMS subsystems, converts them into their physical equivalent, and actually accesses the database and data dictionary as they exist on a storage device. Again, the distinction between logical and physical is important in a database environment. The *physical view* of information deals with how information is physically arranged, stored, and accessed on some type of storage device such as a hard disk. The *logical view* of information, on the other hand, focuses on how you as a knowledge worker need to arrange and access information to meet your particular business needs.

Databases and DBMSs provide two really great advantages in separating the logical from the physical view of information. First, the DBMS handles the physical tasks. So you, as a database user, can concentrate solely on your logical information needs. Second, although there is only one physical view of information, there may be numerous knowledge workers who have different logical views of the information in a database.

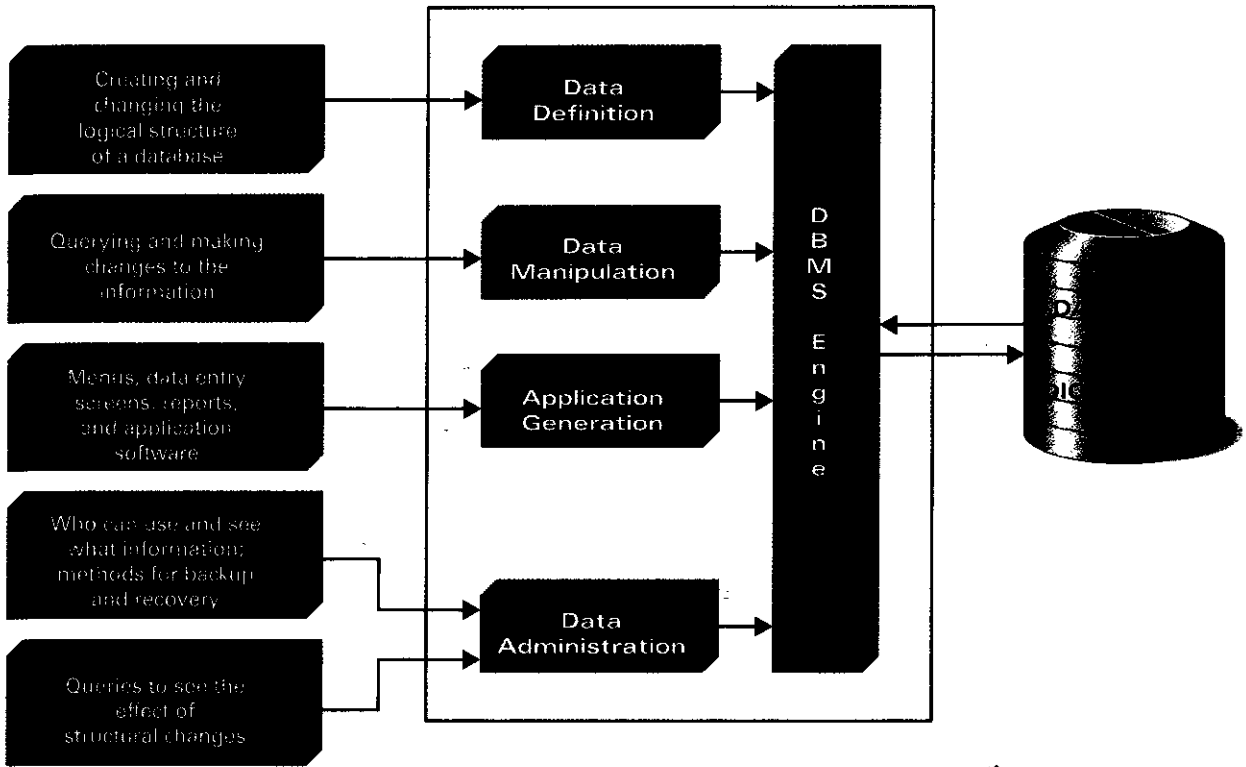


Figure 3.4

Software Subsystems of a Database Management System

That is, according to what business tasks they need to perform, different knowledge workers logically view information in different ways. The DBMS engine can process virtually any logical information view or request into its physical equivalent.

DATA DEFINITION SUBSYSTEM

The *data definition subsystem* of a DBMS helps you create and maintain the data dictionary and define the structure of the files in a database.

When you create a database, you must first use the data definition subsystem to create the data dictionary and define the structure of the files. This is very different from using something like spreadsheet software. When you create a workbook, you can immediately begin typing in information and creating formulas and functions. You can't do that with a database. You must define its logical structure before you can begin typing in any information. Typing in the information is the easy part: Defining the logical structure is more difficult. In *Extended Learning Module C* that follows this chapter, we take you through the process of defining the logical structure for the supply chain management (SCM) side of Solomon Enterprises' database. We definitely recommend that you read that module—knowing how to define the correct structure of a database can be a substantial career opportunity for you.

If you ever find that a certain file needs another piece of information, you have to use the data definition subsystem to add a new field in the data dictionary. Likewise, if you want to delete a given field for all the records in a file, you must use the data definition subsystem to do so.

As you create the data dictionary, you're essentially defining the logical properties of the information that the database will contain. Logical structures of information include the following:

Logical Properties	Examples
Field name	<i>Customer Number, Order Date</i>
Type	Alphabetic, numeric, date, time, etc.
Form	Is an area code required for a phone number?
Default value	If no <i>Order Date</i> is entered, the default is today's date.
Validation rule	Can <i>Amount</i> exceed 8?
Is an entry required?	Must you enter <i>Delivery Address</i> for an order or can it be blank?
Can there be duplicates?	Primary keys cannot be duplicates; but what about amounts?

These are all important logical properties to a lesser or greater extent depending on the type of information you're describing. For example, a typical concrete delivery truck can hold at most eight cubic yards of concrete. Further, Solomon may not accept orders for less than four cubic yards of concrete. Therefore, an important validation rule for *Amount* in the *Order* file is "must be greater than or equal to 4 and cannot be greater than 8."

DATA MANIPULATION SUBSYSTEM

The *data manipulation subsystem* of a DBMS helps you add, change, and delete information in a database and query it for valuable information. Software tools within the data manipulation subsystem are most often the primary interface between you as a user and the information contained in a database. So, while the DBMS engine handles your information requests from a physical point of view, it is the data manipulation tools within a DBMS that allow you to specify your logical information requirements. Those logical information requirements are then used by the DBMS engine to access the information you need from a physical point of view.

In most DBMSs, you'll find a variety of data manipulation tools, including views, report generators, query-by-example tools, and structured query language.

VIEWS A *view* allows you to see the contents of a database file, make whatever changes you want, perform simple sorting, and query to find the location of specific information. Views essentially provide each file in the form of a spreadsheet workbook. The screen in Figure 3.5 shows a view in Microsoft Access for the *Order* file in Solomon's database. At this point, you can click on any specific field and change its contents. You could also point at an entire record and click on the Cut icon (the scissors) to remove a record. If you want to add a record, simply click in the *Order Number* field of the first blank record and begin typing.

Note: we've sorted the file in ascending order by *Concrete Type*. You can easily achieve this by clicking on the A→Z Sort button in the view window. If you want to sort in descending order by *Concrete Type*, simply point to any *Concrete Type* field and click on the Z→A Sort button. You can also perform searches within views. For example, if you wanted to find all orders for *Customer Number* 3456, simply point anywhere in that column, click on the Find Text button (the binoculars), and enter 3456. Access will respond by highlighting each *Customer Number* field where 3456 appears.

As with most other types of personal productivity software, DBMSs support such functions and tasks as cutting and pasting, formatting (for example, bolding a field), spell checking, hiding columns (just as you would do using spreadsheet software), filtering, and even adding links to Web sites.

Sort using these two buttons

Find information with the binoculars

Order Number	Order Date	Customer Number	Delivery Address	Concrete Type	Amount	Truck Number	Driver ID
100000	9/1/2004	1234	55 Smith Lane	1	6	111	123456789
100001	9/1/2004	3456	2122 E. Biscayne	1	3	222	789034444
100005	9/4/2004	5678	1222 Westminster	1	4	222	789034444
100006	9/5/2004	1234	222 East Hampton	1	4	111	123456789
100008	9/8/2004	5789	4532 Lane Circle	1	8	222	789034444
100014	9/10/2004	8789	4225 W. Lawn	1	6	222	984588788
100015	9/9/2004	4567	1333 Burr Ridge	2	4	333	435298657
100016	9/4/2004	4567	1333 Burr Ridge	2	8	222	789034444
100017	9/8/2004	2345	9 W. Palm Beach	2	5	333	789034444
100018	9/9/2004	5789	4532 Lane Circle	2	7	222	435298657
100012	9/8/2004	2345	1111 City Park	2	8	333	435298657
100009	9/7/2004	1234	987 Furlong	3	8	111	123456789
100013	9/10/2004	4567	3500 Tomahawk	4	4	111	789034444
100015	9/12/2004	1234	13 Garden Way	4	8	222	435298657
100002	9/2/2004	1234	55 Smith Lane	5	6	222	435298657
100011	9/8/2004	4567	3500 Tomahawk	5	6	222	789034444
0		0		0	0	0	0

Click here to enter a new record

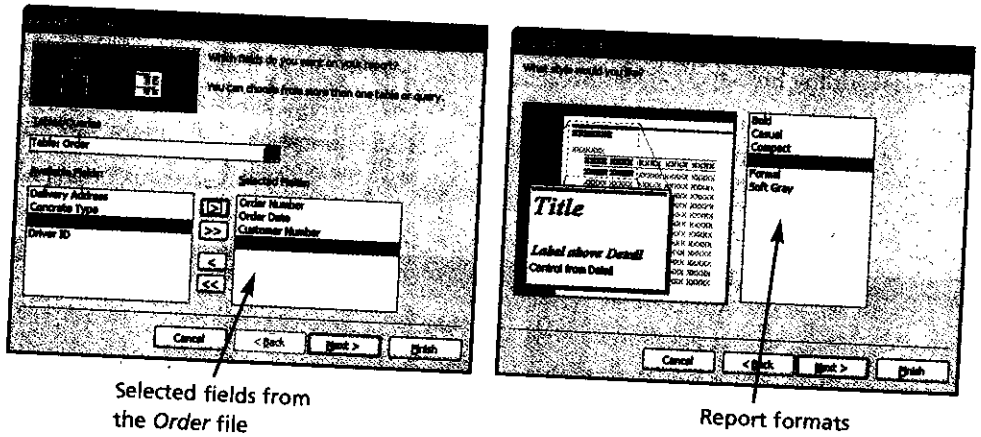
Figure 3.5

A View in Microsoft Access

REPORT GENERATORS *Report generators* help you quickly define formats of reports and what information you want to see in a report. Once you define a report, you can view it on the screen or print it. Figure 3.6 shows two intermediate screens in Microsoft Access. The first allows you to specify which fields of information are to appear in a report. We have chosen to include *Order Number*, *Order Date*, *Customer Number*, and *Amount* from the *Order* file. The second allows you to choose from a set of predefined report formats. Following a simple and easy-to-use set of screens (including the two in Figure 3.6), we went on to specify that sorting should take place by *Customer Number* and that the name of the report should be “Customer and Amount Report.” The completed report is also shown in Figure 3.6. Notice that it displays only those fields we requested, that it’s sorted by *Customer Number*, and that the title is “Customer and Amount Report.”

A nice feature about report generators is that you can save a report format that you use frequently. For example, if you think you’ll use the report in Figure 3.6 often, you can save it by giving it a unique name. Later, you can request that report and your DBMS will generate it, using the most up-to-date information in the database. You can also choose from a variety of report formats (we chose a simple one for our illustration). And you can choose report formats that create intermediate subtotals and grand totals, which can include counts, sums, averages, and the like.

Figure 3.6
Using a Report Generator



Selected fields from the *Order* file

Report formats

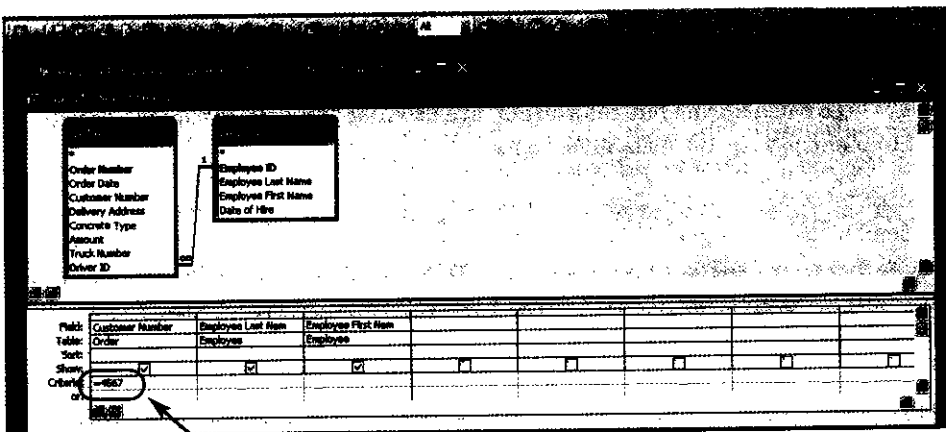
CUSTOMER AND AMOUNT REPORT			
Customer Number	Order Number	Order Date	Amount
	100000		8
1234	100002	9/2/2004	6
	100006		4
1234	100009	9/7/2004	8
	100015		8
2345	100007	9/6/2004	5
	100012		8
3456	100001	9/1/2004	3
	100003		4
4567	100004	9/4/2004	8
	100011		6
4567	100013	9/10/2004	4
	100005		4
6789	100008	9/6/2004	8
	100010		7
6789	100014	9/10/2004	6

QUERY-BY-EXAMPLE TOOLS *Query-by-example (QBE) tools* help you graphically design the answer to a question. Suppose for example that Janielle Smith from Triple A Homes (*Customer Number 4567*) has called and ordered a delivery of concrete. Although she can't remember the name of the driver, she would like to have the driver that comes out the most often to deliver concrete to Triple A Homes. Solomon's task, from a customer relationship management point of view, is to go through all the orders and determine which employee most often delivers concrete to Triple A Homes. The task may seem simple considering that Solomon currently has very few orders in its database. However, can you imagine trying to answer that question if there were thousands of orders in Solomon's database? It would not be fun.

Fortunately, QBE tools can help you answer this question and perform many other queries in a matter of seconds. In Figure 3.7, you can see a QBE screen that formulates the answer to the question. When you perform a QBE, you (1) identify the files in which the needed information is located, (2) drag any necessary fields from the identified files to the QBE grid, and (3) specify selection criteria.

For the names of employees who have delivered concrete to Triple A Homes, we identified the two files of *Order* and *Employee*. Second, we dragged *Customer Number* from the *Order* file to the QBE grid and dragged *Employee Last Name* and *Employee First Name* from the *Employee* file to the QBE grid. Finally, we specified in the Criteria box that we wanted to view only the orders for *Customer Number 4567* (Triple A Homes). Access did the rest and provided the information in Figure 3.7.

QBEs rely heavily on the logical relationships within a database to find information. For example, *Order Number 100004* has the *Customer Number* of 4567 (Triple A Homes). So, the QBE tool took the *Driver ID* from the *Order* file for that order and found a match in the *Employee* file. When it found a match, it presented the *Employee Last Name* and *Employee First Name* (John Robertson). Without the logical relationships being correctly defined, this QBE query would not have worked properly.



Our selection criteria

Customer Number	Employee Last Name	Employee First Name
	Evaraz	
4567	Robertson	John
	Robertson	
4567	Robertson	John

Figure 3.7

Using a Query-by-Example to Find Information

STRUCTURED QUERY LANGUAGE *Structured query language (SQL)* is a standardized fourth-generation query language found in most DBMSs. SQL performs the same function as QBE, except that you perform the query by creating a statement instead of pointing, clicking, and dragging. The basic form of an SQL statement is

```
SELECT . . . FROM . . . WHERE . . .
```

After the **SELECT**, you list the fields of information you want; after the **FROM**, you specify what logical relationships to use; and after the **WHERE**, you specify any selection criteria. If you consider our QBE above of Who most often delivers concrete to *Customer Number 4567*? the SQL statement would look like the following:

```
SELECT Order.[Customer Number], Employee.[Employee Last
Name], Employee.[Employee First Name]
FROM Employee INNER JOIN [Order] ON Employee.[Employee ID]
= Order.[Driver ID]
WHERE (((Order.[Customer Number])=4567));
```

Thoroughly introducing you to the syntax of building SQL statements is outside the scope of this text and would easily require almost 100 pages of material. But you should be aware that SQL does exist. If you're majoring in IT or MIS, you'll undoubtedly take a course in SQL.

APPLICATION GENERATION SUBSYSTEM

The *application generation subsystem* of a DBMS contains facilities to help you develop transaction-intensive applications. These types of applications usually require that you perform a detailed series of tasks to process a transaction. Application generation subsystem facilities include tools for creating visually appealing and easy-to-use data entry screens, programming languages specific to a particular DBMS, and interfaces to commonly used programming languages that are independent of any DBMS.

As with SQL, application generation facilities are most often used by IT specialists. As a knowledge worker, we recommend that you leave application generation to IT specialists as much as you can. You need to focus on views, report generators, and QBE tools. These will help you find information in a database and perform queries so you can start to build and use business intelligence.

DATA ADMINISTRATION SUBSYSTEM

The *data administration subsystem* of a DBMS helps you manage the overall database environment by providing facilities for backup and recovery, security management, query optimization, concurrency control, and change management. The data administration subsystem is most often used by a data administrator or database administrator—someone responsible for assuring that the database (and data warehouse) environment meets the entire information needs of an organization.

Backup and recovery facilities provide a way for you to (1) periodically back up information contained in a database and (2) restart or recover a database and its information in case of a failure. These are important functions you cannot ignore in today's information-based environment. Organizations that understand the importance of their information take precautions to preserve it, often by running backup databases, a DBMS, and storage facilities parallel to the primary database environment. In Chapters 7 and 8, we talk specifically about how to develop plans and strategies in the event of some sort of failure. We call this contingency planning or disaster recovery planning.

LUFTHANSA WANTS YOUR COMPLAINTS

In a time when all airlines are seeking ways to gain passengers and their loyalty, Lufthansa is taking an approach like no other. Simply put, if you don't like something about your flight—the late departure/arrival, uncomfortable seats, hot meals that weren't hot, or anything else you care to complain about—Lufthansa wants to know about it.

This approach isn't to let you vent so that you'll feel better. Lufthansa takes every complaint and enters it into a database. The database, Oracle 9i, supports Lufthansa's COSMIC project—Customer Oriented Service Management Improvement in the Cabin. In any given month, Lufthansa employees enter 6,000 to 7,000 complaints about arrival/departure, boarding, meals, and other aspects of in-flight service. The system tracks such a level of detail that a complaint can even be recorded for a portion of food (e.g., your bread or meat) that you didn't feel was sufficient in size.

That many complaints may seem high, but consider that Lufthansa uses 14,000 cabin attendants to fly 45 million passengers each year to 350 different destinations in 94 countries. That means Lufthansa receives and records only one complaint per 450 passengers. That's good, but not good enough for Lufthansa.

Lufthansa uses Oracle's Discoverer data-mining tools to carefully sift through all the complaints. Lufthansa can easily categorize complaints by each catering caterer. It uses that information to impose penalties on caterers who continually provide substandard food and beverages.

The Discoverer tool set has allowed Lufthansa to cut the time spent handling customer complaints by an amazing 70 percent. That not only saves money, but also increases customer loyalty and retention.

Soon, Lufthansa plans to create satellite links between its in-flight planes and the home office, which will allow cabin attendants to enter complaints on board a flight and have the complaints immediately analyzed.

Customer service is about providing exceptional service. But it's also about rectifying customer complaints. Using databases and tool sets that allow you to sift through and organize customer complaints, your organization can gain a competitive advantage. Just like Lufthansa.

Security management facilities allow you to control who has access to what information and what type of access those people have. In many database environments, for example, some people may need only view access to database information, but not change privileges. Still others may need the ability to add, change, and/or delete information in a database. Through a system of passwords and access privileges, the data administration subsystem allows you to define which users can perform which tasks and what information they can see. At car dealership JM Family Enterprises (JMFE), security management facilities are an absolute must because its technology is highly decentralized and includes users of mobile technologies.⁶ JMFE's system supports encryption and passwords to protect databases, files, and many hardware resources. The system even supports automatic log-offs after a certain amount of time if users accidentally leave their systems running.

Query optimization facilities often take queries from users (in the form of SQL statements or QBEs) and restructure them to minimize response times. In SQL, for example, you can build a query statement that might involve working with as many as 10 different files. As you might well guess, when working with 10 different files, there may be several different solutions for combining them to get the information you need. Fortunately, you don't have to worry about structuring the SQL statement in the most optimized fashion. The query optimization facilities will do that for you and provide you with the information you need in the fastest possible way.

Reorganization facilities continually maintain statistics concerning how the DBMS engine physically accesses information. In maintaining those statistics, reorganization facilities can optimize the physical structure of a database to further increase speed and performance. For example, if you frequently access a certain file by a specific order, the reorganization facilities may maintain the file in that presorted order by creating an index that maintains the sorted order in that file. What's really nice is that you don't have to be aware of the changes to your database with respect to physical locations—the DBMS engine will take care of it for you.

Concurrency control facilities ensure the validity of database updates when multiple users attempt to access and change the same information. This is crucial in today's networked business environment. Consider your school's online registration system. What if you and another student try to register for a class with only one seat remaining at exactly the same time? Who gets enrolled in the class? What happens to the person who does not get his or her desired class schedule? These are important questions that must be answered according to your business rules and, once answered, defined in the database environment using concurrency control facilities.

Change management facilities allow you to assess the impact of proposed structural changes to a database environment. For example, if you decide to add a character identifier to a numeric truck number, you can use the change management facilities to see how many files will be affected. Recall that *Truck Number* would be the primary key for a *Truck* file and that it would also be a foreign key in many other files. Sometimes, structural changes may not have much effect on the database (adding a four-digit zip code extension), but others can cause widespread changes that you must assess carefully before implementing.

All these—backup and recovery, security management, query optimization, reorganization, concurrency control, and change management—are vitally important facilities in any DBMS and thus any database environment. As a user and knowledge worker, you probably won't deal with these facilities specifically as far as setting them up and maintaining them is concerned. But how they're set up and maintained will affect what you can do. So knowing that they do exist and understanding their purposes are important.

Data Warehouses and Data Mining

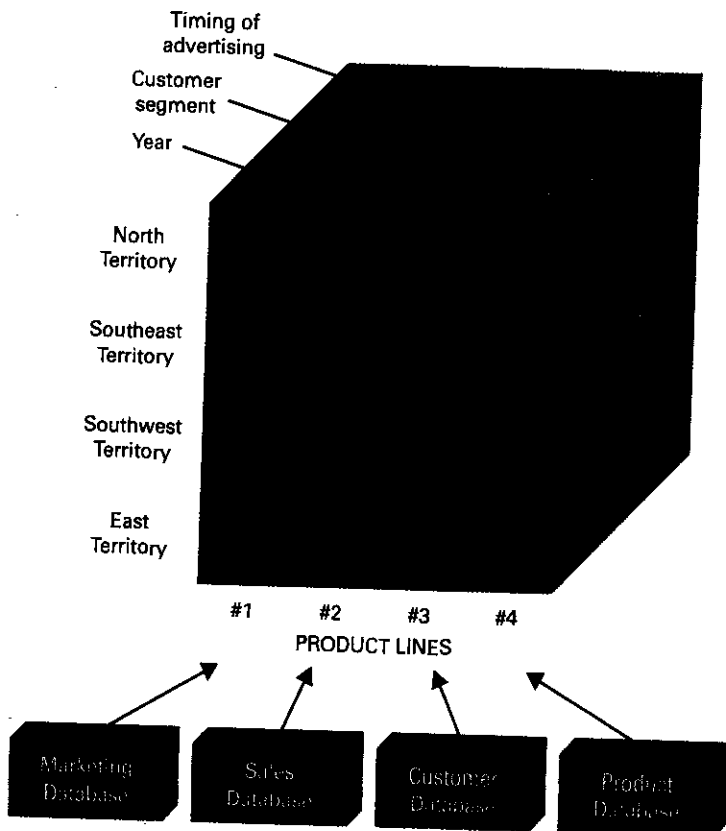
Suppose as a manager at Victoria's Secret, you wanted to know the total revenues generated from the sale of shoes last month. That's a simple query, which you could easily implement using either SQL or a QBE tool. But what if you wanted to know, "By actual versus budgeted, how many size 8 shoes in black did we sell last month in the southeast and southwest regions, compared with the same month over the last 5 years?" That task seems almost impossible, even with the aid of technology. If you were actually able to build a QBE query for it, you would probably bring the organization's operational database environment to its knees.

This example illustrates the two primary reasons so many organizations are opting to build data warehouses. First, while operational databases may have the needed information, the information is not organized in a way that lends itself to building business intelligence within the database or using various data manipulation tools. Second, if you could build such a query, your operational databases, which are probably already supporting the processing of hundreds of transactions per second, would seriously suffer in performance when you hit the Start button to perform the query.

To support such intriguing, necessary, and complex queries to create business intelligence, many organizations are building data warehouses and providing data-mining tools. A data warehouse is simply the next step (beyond databases) in the progression of building business intelligence. And data-mining tools are the tools you use to mine a data warehouse and extrapolate the business intelligence you need to make a decision, solve a problem, or capitalize on an opportunity to create a competitive advantage.

Figure 3.8

A Multidimensional Data Warehouse with Information from Multiple Operational Databases



WHAT IS A DATA WAREHOUSE?

A *data warehouse* is a logical collection of information—gathered from many different operational databases—used to create business intelligence that supports business analysis activities and decision-making tasks (see Figure 3.8). Sounds simple enough on the surface, but data warehouses represent a fundamentally different way of thinking about organizing and managing information in an organization. Consider these key features of a data warehouse, detailed in the sections that follow.

DATA WAREHOUSES ARE MULTIDIMENSIONAL In the relational database model, information is represented in a series of two-dimensional files or tables. Not so in a data warehouse—most data warehouses are multidimensional, meaning that they contain layers of columns and rows. For this reason, most data warehouses are really *multidimensional databases*. The layers in a data warehouse represent information according to different dimensions. This multidimensional representation of information is referred to as a *hypercube*.

In Figure 3.8, you can see a hypercube that represents product information by product line and region (columns and rows), by year (the first layer), by customer segment (the second layer), and by the timing of advertising media (the third layer). Using this hypercube, you can easily ask, According to customer segment A, what percentage of total sales for product line 1 in the southwest territory occurred immediately after a radio advertising blitz? The information you would receive from that query constitutes business intelligence.

Any specific subcube within the larger hypercube can contain a variety of summarized information gathered from the various operational databases. For example, the forward-most and top-left subcube contains information for the North territory, by year, for product line 1. So, it could contain totals, average, counts, and distributions summarizing in some way that information. Of course, what it contains is really up to you and your needs.

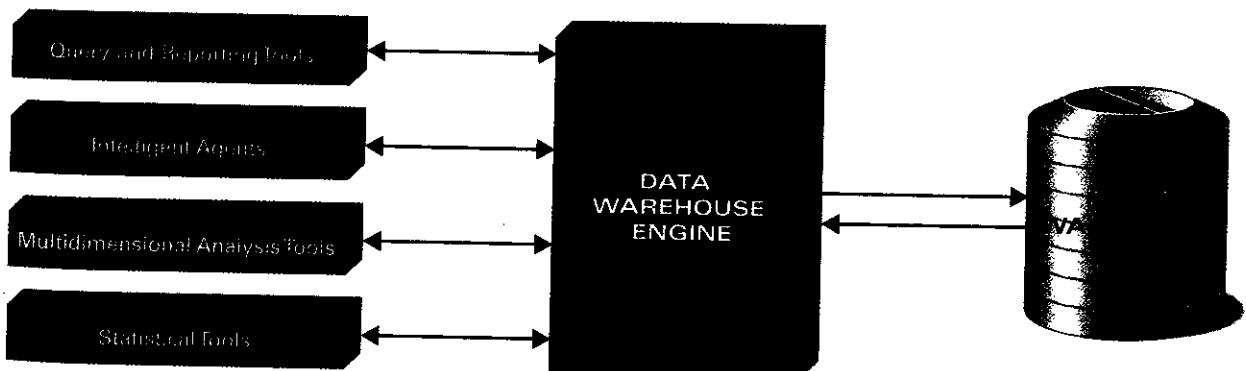
DATA WAREHOUSES SUPPORT DECISION MAKING, NOT TRANSACTION PROCESSING In an organization, most databases are transaction-oriented. That is, most databases support online transaction processing (OLTP) and, therefore, are operational databases. Data warehouses are not transaction-oriented: They exist to support decision-making tasks in your organization. Therefore, data warehouses support only online analytical processing (OLAP).

As we just stated, the subcubes within a data warehouse contain summarized information. So, while a data warehouse may contain the total sales for a year by product line, it does not contain a list of each individual sale to each individual customer for a given product line. Therefore, you simply cannot process transactions with a data warehouse. Instead, you process transactions with your operational databases and then use the information contained within the operational databases to build the summary information in a data warehouse.

WHAT ARE DATA-MINING TOOLS?

Data-mining tools are the software tools you use to query information in a data warehouse. These data-mining tools support the concept of OLAP—the manipulation of information to support decision-making tasks. Data-mining tools include query-and-reporting tools, intelligent agents, multidimensional analysis tools, and statistical tools (see Figure 3.9). Essentially, data-mining tools are to data warehouse users what data manipulation subsystem tools are to database users.

Figure 3.9
The Data Miner's Tool Set



As in a DBMS, a data warehouse system has an engine responsible for converting your logical requests into their physical equivalent.

A PERFECT MATCH—DATA WAREHOUSES, BUSINESS INTELLIGENCE, AND CUSTOMER RELATIONSHIP MANAGEMENT

Bank Hapoalim is Israel's largest bank, serving almost 2 million customers with a wide array of financial services and products. Bank Hapoalim's customer relationship management (CRM) focus is to offer superior customer care and unique financial services to maintain and expand its customer base.

As a part of this effort, Bank Hapoalim turned to Cognos, the world leader in business intelligence software and data warehouses. According to Tal Shlasky, data warehouse project manager at Bank Hapoalim, "We continue to devote our resources and energy to

the individual customer in our quest to be more efficient and better suited to banking in the 21st century. Using an industry-leading solution like Cognos, we are doing just that. Cognos gives us an integrated view of the enterprise, allowing us to target customers based on their individual needs. Without a doubt, that capability gives us a competitive edge in the market."

To grow a customer base of 2 million, any organization needs data warehouses that support business intelligence.

QUERY-AND-REPORTING TOOLS *Query-and-reporting tools* are similar to QBE tools, SQL, and report generators in the typical database environment. In fact, most data warehousing environments support simple and easy-to-use data manipulation subsystem tools such as QBE, SQL, and report generators. Most often, data warehouse users use these types of tools to generate simple queries and reports.

INTELLIGENT AGENTS Intelligent agents utilize various artificial intelligence tools such as neural networks and fuzzy logic to form the basis of "information discovery" and building business intelligence in OLAP. For example, Wall Street analyst Murray Riggiero uses OLAP software called Data/Logic, which incorporates neural networks to generate rules for his highly successful stock and bond trading system.⁹ Other OLAP tools, such as Data Engine, incorporate fuzzy logic to analyze real-time technical processes.

Intelligent agents represent the growing convergence of various IT tools for working with information. Previously, intelligent agents were considered only within the context of artificial intelligence and were seldom thought to be a part of the data organizing and managing functions in an organization. Today, you can find intelligent agents being used not only for OLAP in a data warehouse environment but also for searching for information on the Web. In Chapter 4, we'll explore artificial intelligence techniques such as intelligent agents.

MULTIDIMENSIONAL ANALYSIS TOOLS *Multidimensional analysis (MDA) tools* are slice-and-dice techniques that allow you to view multidimensional information from different perspectives. For example, if you completed any of the recommended group projects for Chapter 1, you were using spreadsheet software to literally slice and dice the provided information. Within the context of a data warehouse, we refer to this process as "turning the cube." That is, you're essentially turning the cube to view information from different perspectives.

This turning of the cube allows you to quickly see information in different subcubes. If you refer back to the data warehouse in Figure 3.8 on page 142, you'll notice that information by customer segment and timing of advertising is actually hidden. Using MDA tools, you can easily bring this to the front of the data warehouse for viewing. What

you've essentially done is to slice the cube vertically by layer and bring some of the background layers to the front. As you do this, the values of the information are not affected.

STATISTICAL TOOLS Statistical tools help you apply various mathematical models to the information stored in a data warehouse to discover new information. For example, you can perform a time-series analysis to project future trends. You can also perform a regression analysis to determine the effect of one variable on another.

Sega of America, one of the largest publishers of video games, uses a data warehouse and statistical tools to effectively distribute its advertising budget of more than \$50 million a year.^{10,11} With its data warehouse, product line specialists and marketing strategists "drill" into trends of each retail store chain. Their goal is to find buying trends that will help them better determine which advertising strategies are working best (and at what time of the year) and how to reallocate advertising resources by media, territory, and time. Sega definitely benefits from its data warehouse, and so do retailers such as Toys "Я" Us, Wal-Mart, and Sears—all good examples of customer relationship management through technology.

To learn more about today's best data warehousing and data-mining tools, visit the Web site that supports this text at www.mhhe.com/haag.

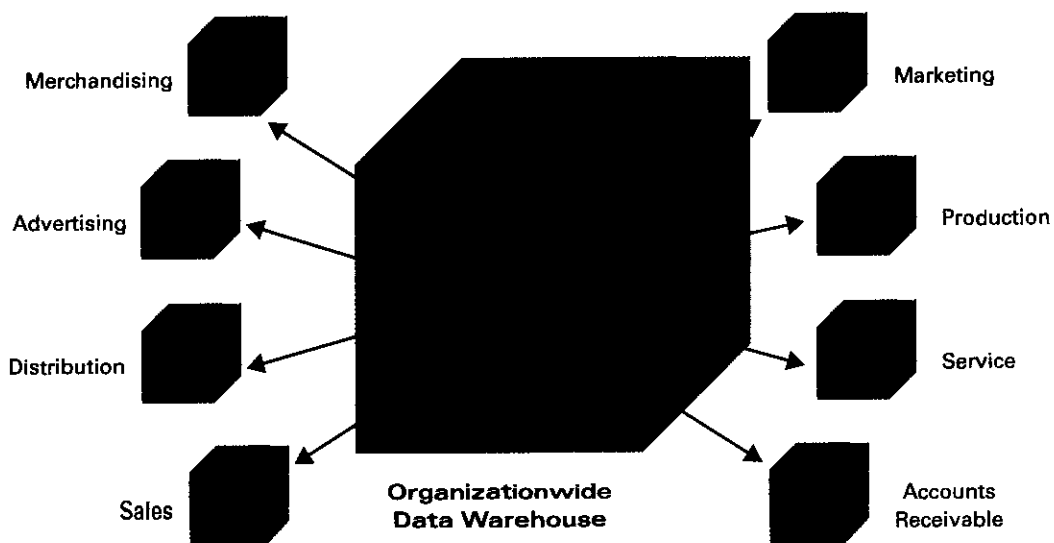
DATA MARTS: SMALLER DATA WAREHOUSES

Data warehouses are often perceived as organizationwide, containing summaries of all the information that an organization tracks. However, some people need access to only a portion of that data warehouse information as opposed to all of it. In this case, an organization can create one or more data marts. A *data mart* is a subset of a data warehouse in which only a focused portion of the data warehouse information is kept (see Figure 3.10).

Lands' End first created an organizationwide data warehouse for everyone to use, but soon found out that there can be "too much of a good thing."¹² In fact, many Lands' End employees wouldn't use the data warehouse because it was simply too big, too complicated, and included information they didn't need access to. So, Lands' End created several smaller data marts. For example, Lands' End created a data mart just for the merchandising department. That data mart contains only merchandising-specific information and not any information, for instance, that would be unique to the finance department.

Figure 3.10

Data Marts Are Subsets of Data Warehouses



Because of the smaller, more manageable data marts, knowledge workers at Lands' End are making better use of information. If some of your employees don't need access to organizationwide data warehouse information, consider building a smaller data mart for their particular needs.

If you do choose to build smaller data marts for your employees, the data-mining tools are the same. That is, data marts support the use of query-and-reporting tools, intelligent agents, multidimensional analysis tools, and statistical tools. This yields efficiency in an organization with respect to training. Once you've trained your employees to use any or all data-mining tools, they can apply them to an organizationwide data warehouse or smaller data marts.

DATA MINING AS A CAREER OPPORTUNITY

Data mining represents a substantial career opportunity for you, no matter what your career choice. In the business world, you'll face numerous situations in which you need business intelligence to make the right and most effective decisions.

Fortunately, you don't have to be an IT expert to perform data mining. As you'll learn in *Extended Learning Module D* (Decision Analysis with Spreadsheet Software), you can actually use a spreadsheet tool such as Microsoft Excel to build a three-dimensional cube similar to the one in Figure 3.8. You can then use Excel's other decision support features to build a graph, perform a regression analysis, and "turn the cube" by bringing new layers of information forward. You can do the same with Microsoft Access, by building a three-dimensional cube (i.e., data warehouse) of information stored in a database. We definitely recommend that you learn to use these tools and then note your proficiency in your e-portfolio under "Technology Skills."

Beyond personal productivity tools, you should consider learning how to use some data-mining tools specific to the data warehouse environment. Some of the more popular ones include:

HOW UP-TO-DATE SHOULD DATA WAREHOUSE INFORMATION BE?

Information timeliness is a must in a data warehouse—old and obsolete information leads to poor decision making. Below is a list of decision-making processes that people go through for different business environments. For each, specify whether the information in the data warehouse should be updated monthly, weekly, daily, or by the minute. Be prepared to justify your decision.

1. To adjust class sizes in a university registration environment
2. To alert people to changes in weather conditions
3. To predict scores for professional football games
4. To adjust radio advertisements in light of demographic changes
5. To monitor the success of a new product line in the clothing retail industry
6. To adjust production levels of foods in a cafeteria
7. To switch jobs to various printers in a network
8. To adjust CD rates in a bank
9. To adjust forecasted demands of tires in an auto parts store

- Query and Analysis and Enterprise Analytic tools in Business Objects (www.businessobjects.com)
- Business Intelligence and Information Access tools in SAS (www.sas.com)
- ReportNet, PowerPlay, Visualizer, NoticeCast, and DecisionStream tools in Cognos (www.cognos.com)
- PowerAnalyzer tools in Informatica (www.informatica.com)

There are many, many others. You should have a look at your school's catalog of courses in data mining—you may find them offered in the technology department, statistics department, and other departments. We recommend that at the very least you become acquainted with the following: SAS (the leading vendor in statistical software), Cognos (the leading vendor in data warehousing and data-mining tools), and Informatica (the second-leading vendor in data warehousing and data-mining tools).

IMPORTANT CONSIDERATIONS IN USING A DATA WAREHOUSE

As is true with all types of technology, you can't simply implement a data warehouse and use data-mining tools just because they're a "hot" set of technologies and expect automatically to increase your efficiency and effectiveness. Always let your business needs drive your technology decisions. You have to need the technology and the technology has to fit your needs. With respect to data warehouse and data-mining tools, consider your answers to the following questions.

1. **Do you need a data warehouse?** Although great IT tools, they are not necessarily the best technologies for all businesses because (1) they are expensive, (2) they may not be necessary since some businesses can easily extract all the business intelligence they need from databases, and (3) they require extensive and often expensive support.
2. **Do all your employees need an entire data warehouse?** If not, consider building data marts.

3. **How up-to-date must the information be?** To create a data warehouse, you take “snapshots” of database information and load it into a data warehouse. If crucial information changes every second, this may not be possible.
4. **What data-mining tools do you need?** User needs should always drive the answer to this question. Whichever you choose, training will be key. If your users can fully exploit all the features of their chosen data-mining tools, your entire organization will reap the benefits.

Managing the Information Resource

As you prepare to enter today’s fast-paced, exciting, and information-based business world, you must be prepared to help your organization manage and organize its information. After all, you will be a knowledge worker—a person who works primarily with information. Your organization will be successful, in part, because of your ability to organize and manage information in a way that best moves the organization toward its goals. We would offer you three important questions to keep in mind. The answers to them are definitely moving targets. As business and technology change, your answers may have to change as well.

WHO SHOULD OVERSEE YOUR ORGANIZATION’S INFORMATION RESOURCE?

Organizations today can have chief executive officers (CEOs), chief operating officers (COOs), and chief financial officers (CFOs), among others. You can also find another title—chief information officer. The *chief information officer (CIO)* is responsible for overseeing an organization’s information resource. A CIO’s responsibilities may range from approving new development activities for data warehouses and data marts to monitoring the quality and use of information within those data warehouses and data marts.

Two important functions associated with overseeing an organization’s information resource are data administration and database administration. *Data administration* is the function in an organization that plans for, oversees the development of, and monitors the information resource. It must be completely in tune with the strategic direction of the organization to assure that all information requirements can and are being met.

Database administration is the function in an organization that is responsible for the more technical and operational aspects of managing the information contained in organizational information repositories (databases, data warehouses, and data marts). Database administration functions include defining and organizing database structures and contents, developing security procedures, developing database and DBMS documentation, approving and monitoring the development of databases and database applications, reporting/rectifying intrusions or operational anomalies, and maintaining overall system integrity.

In large organizations, both administrative functions are usually handled by steering committees rather than by a single individual. These steering committees are responsible for their respective functions and for reporting to the CIO. It’s definitely a team effort to manage most organizational resources—information is no different from other assets in that it needs careful oversight and management.

However, in certain ways information is different from many “typical” organizational resources. For example, information is intangible, so it becomes extremely difficult to measure its worth. What dollar value would you attach to a customer record? Realistically, it is impossible to set a precise dollar value on information. So, how can you know

how much you should spend on information technology tools? It is a difficult question to answer.

Since information is intangible, it can also be shared by numerous people and not actually be “consumed.” Money is different. If you have money in your department budget and spend it on travel, you can’t very well use that same money for employee education expenses. In this way, then, the intangibility of information is good because many people can use it without “consuming” it. But as we’ve alluded to many times, this is also why you have to make special and unique considerations for the security of information that you do not have to make for other organizational resources, such as money. Security is mostly the responsibility of database administration.

IS INFORMATION OWNERSHIP A CONSIDERATION?

Information sharing in your organization means that anyone—regardless of title or department—can access and use whatever information he or she needs. But information sharing brings to light an important question: Does anyone in your organization own any information? In other words, if everyone shares information, who is ultimately responsible for providing that information and assuring the quality of the information? Information ownership is a key consideration in today’s information-based business environment. Someone must accept full responsibility for providing specific pieces of information and ensuring the quality of that information. If you find that the wrong information is stored in the organization’s data warehouse, you must be able to determine the source of the problem and whose responsibility it is.

This issue of information ownership is similar to other management functions. If you manage a department, you’re responsible for the work in that department as well as its expenses and people. The same is true for information. If information originates in your department, you essentially own that information because you’re providing it to those who need it and ensuring its quality.

HOW “CLEAN” MUST YOUR INFORMATION BE?

Information “cleanliness” (closely related to information ownership) is an important topic today and will be for many years. Have you ever received the same piece of advertising mail (snail mail, that is) multiple times from the same company on the same day?

Many people have, and it's an example of "unclean" information. In this instance, your name may appear twice in a database, once without your middle initial and once with your middle initial. Likewise, your name may appear twice in a database with two different spellings of your last name.

In all popular business-oriented databases and DBMSs, such as Oracle, you can find utilities to help you "clean" your information. In the example above of having your information twice in a database (with different spellings of your last name), the utility would probably determine that the two records actually belong to the same person (you) because of the identical nature of other associated information such as your address and phone number.

In a data warehouse environment, recall that information comes from multiple databases. It becomes even more important then to consolidate and "clean" the information so no duplicate or erroneous information exists in the data warehouse. To accomplish this, you go through *extraction, transformation, and loading (ETL)*. ETL is a process in which you specify what information you want from each database, how that information is to be associated, and what rules to follow in consolidating the information to ensure its cleanliness. You certainly don't want duplicate information weakening the foundation for your business intelligence.

Summary: Student Learning Outcomes Revisited

- 1. Describe business intelligence and its role in an organization.** *Business intelligence* is knowledge—knowledge about your customers, your competitors, your partners, your competitive environment, and your own internal operations—that gives you the ability to make effective, important, and often strategic business decisions. Business intelligence is much more than just a list of your products or to whom you've sold them. It would combine your product information perhaps with your advertising strategy information and customer demographics to help you determine the effectiveness of various advertising media on demographic groups segmented by location.
- 2. Differentiate between databases and data warehouses with respect to their focus on online transaction processing and online analytical processing.** A *database* is a collection of information that you organize and access according to the logical structure of that information. Databases support both online

transaction processing (OLTP) and online analytical processing (OLAP). Databases that support OLTP are often referred to as **operational databases**. These databases contain detailed information about transactions that have taken place. And using various data manipulation tools, you can query a database to extract meaningful information. A **data warehouse** is a collection of information—gathered from many different operational databases—used to create business intelligence that supports business analysis activities and decision-making tasks. So, data warehouses support only OLAP, not OLTP.

3. **List and describe the key characteristics of a relational database.** The **relational database model** uses a series of logically related two-dimensional tables or files to store information in the form of a database. Key characteristics include

- A collection of information—Composed of many files or tables of information that are related to each other
- Contain logical structures—You care only about the logical information and not about how it's physically stored or where it's physically located
- Have logical ties among the information—All the files in a database are related in that some primary keys of certain files appear as foreign keys in others
- Possess built-in integrity constraints—When creating the data dictionary for a database, you can specify rules by which the information must be entered (e.g., not blank, etc.)

4. **Define the five software components of a database management system.** The five software components of a database management system include

- **DBMS engine**—Accepts logical requests from the various other DBMS subsystems, converts them into their physical equivalent, and actually accesses the database and data dictionary as they exist on a storage device
- **Data definition subsystem**—Helps you create and maintain the data dictionary and define the structure of the files in a database
- **Data manipulation subsystem**—Helps you add, change, and delete information in a database and query it for valuable information

- **Application generation subsystem**—Contains facilities to help you develop transaction-intensive applications
- **Data administration subsystem**—Helps you manage the overall database environment by providing facilities for backup and recovery, security management, query optimization, concurrency control, and change management

5. **List and describe the key characteristics of a data warehouse.** The key characteristics of a data warehouse include

- **Multidimensional**—While databases store information in two-dimensional tables, data warehouses include layers to represent information according to different dimensions
- **Support decision making**—Data warehouses, because they contain summarized information, support business activities and decision-making tasks, not transaction processing

6. **Define the four major types of data-mining tools in a data warehouse environment.**

The four major types of data-mining tools in a data warehouse environment include

- **Query-and-reporting tools**—Similar to QBE tools, SQL, and report generators in the typical database environment
- **Intelligent agents**—Utilize various artificial intelligence tools such as neural networks and fuzzy logic to form the basis of “information discovery” and building business intelligence in OLAP
- **Multidimensional analysis (MDA) tools**—Slice-and-dice techniques that allow you to view multidimensional information from different perspectives
- **Statistical tools**—Help you apply various mathematical models to the information stored in a data warehouse to discover new information

7. **List key considerations in managing the information resource in an organization.**

Key considerations in managing the information resource in an organization include these questions:

- Who should oversee the organization's information?
- Is information ownership a consideration?
- How “clean” must your information be?

CLOSING CASE STUDY ONE

BEN & JERRY'S, STAPLES, AND BUSINESS INTELLIGENCE

Organizations want information. Organizations need information. However, information must be in an organized format that supports the creation of business intelligence. Otherwise, according to Rebecca Wettemann, Vice President of Research at Nucleus Research, "It's like having a bank account with millions of dollars in it but no ATM card. If you can't get it [business intelligence] and can't make it work for you, then it is not really useful."

In support of creating and using business intelligence, companies have focused much of their spending efforts on business intelligence software and data-mining tools. According to a Merrill Lynch survey in 2003, business intelligence software and data-mining tools were at the top of the technology spending list of CIOs. And according to A. G. Edwards, the market for that type of software is expected to grow from \$4.7 billion in 2003 to \$7.5 billion in 2006.

Consider two companies—Ben & Jerry's and Staples—and their approach to creating and using business intelligence.

BEN & JERRY'S

Ben & Jerry's, located in Waterbury, Vermont, produces 190,000 pints of ice cream and frozen yogurt daily and ships to over 50,000 grocery stores in the United States and 12 other countries. Every single pint is meticulously tracked, first by being entered into an Oracle database. With that information carefully organized, Ben & Jerry's uses a sophisticated data-mining tool set from a company called Business Objects.

For example, the sales people can easily monitor sales to determine how much ground Cherry Garcia Frozen Yogurt is gaining on Cherry Garcia Ice Cream, its number one selling product. The consumer affairs staff can even correlate each of the several hundred calls and e-mails received each week to the exact pint of ice cream. If complaints are consistent concerning a specific batch, the consumer affairs staff can drill down to the supplier who provided the ingredients such as milk or eggs.

In one particular instance, Ben & Jerry's received a large number of complaints that its Cherry Garcia Ice Cream didn't have enough cherries. The complaints were coming in from all over the country, so it wasn't a regional problem. Employees continued drilling through business intelligence with Business Objects and determined that the manufacturing process (from the supplies of raw materials to the mixing) was satisfactory and had no anomalies. Eventually the problem was determined to be that the ice cream box for Cherry Garcia Ice Cream had on it a photo of frozen yogurt, a product with more cherries than the ice cream. Simply changing the photo on the box solved the problem.

STAPLES

Staples, a \$10.7-billion-a-year office-supply chain, extensively uses statistical tools on its data warehouse information to create business intelligence. Alan Gordon, Director of Sales Forecasting, joined the company in 1993 when the company had 150 stores. His charge was to determine where to build new stores. Using a variety of programs from SAS (the leading provider of statistical tools within business intelligence software), Alan created a sophisticated system that evaluates 40 variables, including proximity to competitors and sales tax by zip code. Of the some 4,000 sites that Alan now targets each year, approximately 100 become a new store. That has translated into 950 new stores since his arrival.

Marci Lerner, Staples' Vice President of Finance, also discovered the true value of statistical data-mining tools and business intelligence. Marci decided to use a suite of statistical tools provided by Hyperion Solutions. Using her new business intelligence, Marci gained valuable insights into not only the company's financial situation but also how to display merchandise in the stores.

For example, Marci determined that Staples had been misusing its floor space. Typically, each store devoted a great deal of space to large items such as file cabinets, desks, and other furniture. It made logical sense because bigger items yielded better gross mar-

gins than pens and pencils. However, Marci's analysis of Staples' business intelligence pointed out that the costs of storage, distribution, handling, damage, and labor associated with large items made them less profitable than smaller, less space-intensive categories of office supplies. Because of that finding, Staples has decreased its furniture department in most stores, in favor of more room for labels, paper, desk organizers, calendars, and the like. This has helped the company grow its net income to the tune of 12 percent compounded over the last five years.^{15,16}

QUESTIONS

1. Ben & Jerry's tracks a wealth of information on each pint of ice cream and frozen yogurt. If you were to design Ben & Jerry's data warehouse, what dimensions of information would you include? As you develop your list of dimensions, consider every facet of Ben & Jerry's business operations, from supply chain management to retail store monitoring.
2. Databases are the underlying technology that allows Ben & Jerry's to track ice cream and frozen yogurt information. Based on your knowledge of databases, what sort of tables or files of information would Ben & Jerry's need in its database? What would be the primary keys for each of those? What would be the foreign keys among those to create the necessary relationships?
3. Marci Lerner, of Staples, used financial business intelligence to determine the best utilization of floor space for product displays and the like. What other business questions and issues can be addressed by looking at financial information within the context of business intelligence? For example, could Marci also use her financial information to determine the optimal distribution lines for moving products to the various stores?
4. Alan Gordon, of Staples, takes into consideration 40 variables to evaluate a potential location for a new store. In this case study, we identified two of those as being proximity to competitors and sales tax by zip code. Make a list of 10 other variables that Alan might also use in his analysis. For each, provide justification.
5. Neil Hastie, CIO at TruServe Corporation, once described most decision making in all types of businesses as "a lot of by-guess and by-golly, a lot of by-gut, and a whole lot of paper reports." That statement is not kind to managers in general or to IT specialists charged with providing the right people with the right technology to make the right decisions. What's the key to turning Neil's statement into a positive one? Is it training? Is it providing timely information access? Is it providing everyone with a wide assortment of data-mining tools? Other solutions? Perhaps it's a combination of several answers.

CLOSING CASE STUDY TWO

MINING DINING DATA

Restaurants, fast-food chains, casinos, and others use data warehouses to determine customer purchasing habits and to determine what products and promotions to offer and when to offer them. Some of the leading data warehouse users include AFC Enterprises (operator and franchiser of more than 3,300 Church's Chicken, Popeyes' Chicken and Biscuits, Seattle Coffee

Company, Cinnabon, and Torrefazione outlets world-wide); Red Robin International (a 170-unit casual-dining chain); Harrah's Entertainment (owner of 26 U.S. casinos); Pizzeria Uno; and Einstein/Noah Bagel (operator of 428 Einstein's and 111 Noah's New York Bagel stores).

AFC ENTERPRISES

AFC Enterprises cultivates a loyal clientele by slicing and dicing its data warehouse to strategically configure promotions and tailor menus to suit local preferences. AFC's data warehouse helps it better understand its core customers and maximize its overall profitability. AFC tracks customer-specific information from name and address to order history and frequency of visits. This enables AFC to determine exactly which customers are likely to respond to a given promotion on a given day of the week.

AFC also uses its data warehouse to anticipate and manipulate customer behavior. For example, AFC can use its data warehouse to determine that coffee is added to the tab 65 percent of the time when a particular dessert is ordered and 85 percent of the time when that dessert is offered as a promotional item. Knowing that, AFC can run more promotions for certain desserts figuring that customers will respond by ordering more desserts and especially more coffee (coffee is a high-margin item in the restaurant business).

RED ROBIN INTERNATIONAL

Red Robin's terabyte-size data warehouse tracks hundreds of thousands of point-of-sale (POS) transactions, involving millions of menu items and more than 1.5 million invoices. As Howard Jenkins, Red Robin's Vice President of Information Systems, explains it, "With data mining in place, we can ask ourselves, 'If we put the items with high margins in the middle of the menu, do we sell more versus putting it at the top or bottom, [and if so], to whom and where?' We can also tell if something cannibalizes the sale of other items and can give the marketing department an almost instant picture of how promotions are being sold and used."

The placement of items on a menu is strategic business, just as the placement of promotional items in a grocery store can mean increased sales for one item and reduced sales for another. The job of finding the right mix is definitely suited to mining a data warehouse.

Using Cognos Business Intelligence, Red Robin now has measurable results of promotion and menu changes, makes better and more timely decisions, and has realized seven-figure savings in operational costs.

HARRAH'S ENTERTAINMENT

Harrah's Entertainment uses its data warehouse to make decisions for its highly successful Total Gold customer recognition program. Depending on their spending records, Total Gold members can receive free vouchers for dining, entertainment, and sleeping ac-

commodations. Knowing which rewards to give to which customers is key.

John Boushy, Senior Vice President of Entertainment and Technology for Harrah's, says, "We can determine what adds value to each customer and provide that value at the right time." Dining vouchers or free tickets for shows are awarded to day visitors, not sleeping accommodations. Customers who consistently visit a particular restaurant and order higher-end foods receive free dinners and cocktails, not vouchers for free (and cheaper) breakfasts.

PIZZERIA UNO

Pizzeria Uno uses its data warehouse to apply the 80/20 rule. That is, it can determine which 20 percent of its customers contribute to 80 percent of its sales and adjust menus and promotions to suit top patron preferences. These changes can often lead to converting some of the other 80 percent of Pizzeria Uno's customers to the more profitable 20 percent.

EINSTEIN/NOAH BAGEL

Einstein/Noah Bagel uses its data warehouse in real time to maximize cross-selling opportunities. For example, if data warehouse information reveals that a manager in a given store might be missing a cross-selling opportunity on a particular day, an e-mail is automatically sent out to alert managers to the opportunity. Salespeople can then respond by offering the cross-selling opportunity ("How about a cup of hot chocolate with that bagel since it's so cold outside?") to the next customer.^{17,18,19}

QUESTIONS

1. Consider the issue of timely information with respect to the businesses discussed in the case. Which of the businesses must have the most up-to-date information in its data warehouse? Which business can have the most out-of-date information in its data warehouse and still be effective? Rank the five businesses discussed with a 1 for the one that needs the most up-to-date information and a 5 for the one that is least sensitive to timeliness of information. Be prepared to justify your rankings.
2. Harrah's Entertainment tracks a wealth of information concerning customer spending habits. If you were to design Harrah's Entertainment's data warehouse, what dimensions of information would you include? As you develop your list of dimensions, consider

- every facet of Harrah's business operations, including hotels, restaurants, and gaming casinos.
3. AFC Enterprises includes information in its data warehouse such as customer name and address. Where does it (or could it) gather such information? Think carefully about this, because customers seldom provide their names and addresses when ordering fast food at a Church's or Popeyes. Is AFC gathering information in an ethical fashion? Why or why not?
 4. Visit a local grocery store and walk down the breakfast cereal aisle. You should notice something very specific about the positioning of the various breakfast cereals. What is it? On the basis of what information do you think grocery stores determine cereal placement? Could they have determined that information from a data warehouse or from some other source? If another source, what might that source be?
 5. Suppose you're opening a pizza parlor in the town where you live. It will be a "take and bake" pizza parlor in which you make pizzas for customers but do not cook them. Customers buy the pizzas uncooked and take them home for baking. You will have no predefined pizza types but will make each pizza to the customer's specifications. What sort of data warehouse would you need to predict the use of toppings by time of day and by day of the week? What would your dimensions of information be? If you wanted to increase the requests for a new topping (such as mandarin oranges), what information would you hope to find in your data warehouse that would enable you to do so?

Key Terms and Concepts

Application generation subsystem, 138
 Business intelligence (BI), 126
 Chief information officer (CIO), 148
 Data administration, 148
 Data administration subsystem, 138
 Database, 128
 Database administration, 148
 Database management system (DBMS), 132
 Data definition subsystem, 133
 Data dictionary, 128
 Data manipulation subsystem, 134
 Data mart, 145
 Data-mining tool, 143
 Data warehouse, 142
 DBMS engine, 132
 Foreign key, 130

Integrity constraint, 131
 Logical view, 132
 Multidimensional analysis (MDA) tool, 144
 Online analytical processing (OLAP), 126
 Online transaction processing (OLTP), 126
 Operational database, 126
 Physical view, 132
 Primary key, 130
 Query-and-reporting tool, 144
 Query-by-example (QBE) tool, 137
 Relation, 128
 Relational database, 128
 Report generator, 136
 Structured query language (SQL), 138
 View, 135

Short-Answer Questions

1. What is business intelligence? Why is it more than just information?
2. What is online transaction processing (OLTP)?
3. What is online analytical processing (OLAP)?
4. What is the most popular database model?
5. How are primary and foreign keys different?
6. What are the five important software components of a database management system?
7. How are QBE tools and SQL similar? How are they different?
8. What is a data warehouse? How does it differ from a database?
9. What are the four major types of data-mining tools?
10. What is a data mart? How is it similar to a data warehouse?
11. What is the role of a chief information officer (CIO)?