Part

I

Contents

Guided Tours

I n Part I, we will tour basic SAP BW (Business Information Warehouse) functionalities using a simplified business scenario—sales analysis.

After introducing the basic concept of data warehousing and giving an overview of BW, we create a data warehouse using BW and load data into it. We then check data quality before creating queries and reports (or workbooks, as they are called in BW). Next, we demonstrate how to use an SAP tool called Profile Generator to manage user authorization.

After finishing the guided tours, we will appreciate BW's ease of use and get ready to explore other BW functionalities.
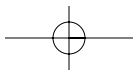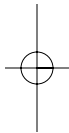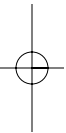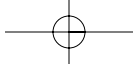
Chapter
1

Business Scenario and SAP BW

The objective of **data warehousing** is to analyze data from diverse sources to support decision making. To achieve this goal, we face two challenges:

- Poor system performance. A data warehouse usually contains a large volume of data. It is not an easy job to retrieve data quickly from the data warehouse for analysis purposes. For this reason, the data warehouse design uses a special technique called a **star schema**.
- Difficulties in extracting, transferring, transforming, and loading (**ETTL**) data from diverse sources into a data warehouse. Data must be cleansed before being used. ETTL has been frequently cited as being responsible for the failures of many data warehousing projects. You would feel the pain if you had ever tried to analyze SAP R/3 data without using SAP BW.

SAP R/3 is an ERP (Enterprise Resources Planning) system that most large companies in the world use to manage their business transactions. Before the introduction of SAP BW in 1997, ETTL of SAP R/3 data into a data warehouse seemed an unthinkable task. This macro-environment explained the urgency with which SAP R/3 customers sought a data warehousing solution. The result is SAP BW from SAP, the developer of SAP R/3.

In this chapter we will introduce the basic concept of data warehousing. We will also discuss what **SAP BW (Business Information Warehouse)** is, explain why we need it, examine its architecture, and define Business Content.

First, we use sales analysis as an example to introduce the basic concept of data warehousing.

1.1 Sales Analysis—A Business Scenario

Suppose that you are a sales manager, who is responsible for planning and implementing sales strategy. Your tasks include the following:

- Monitoring and forecasting sales demands and pricing trends
- Managing sales objectives and coordinating the sales force and distributors
- Reviewing the sales activities of each representative, office, and region

Suppose also that you have the data in Tables 1.1 through 1.3 available about your firm's materials, customers, and sales organization.

TABLE 1.1 MATERIALS	Material Number	Material Name	Material Description
	MAT001	TEA	Ice tea
	MAT002	COFFEE	Hot coffee
	MAT003	COOKIE	Fortune cookie
	MAT004	DESK	Computer desk
	MAT005	TABLE	Dining table
	MAT006	CHAIR	Leather chair
	MAT007	BENCH	Wood bench
	MAT008	PEN	Black pen
	MAT009	PAPER	White paper
	MAT010	CORN	America corn
	MAT011	RICE	Asia rice
	MAT012	APPLE	New York apple
	MAT013	GRAPEFRUIT	Florida grapefruit
	MAT014	PEACH	Washington peach
	MAT015	ORANGE	California orange

TABLE 1.2
CUSTOMERS

Customer ID	Customer Name	Customer Address
CUST001	Reliable Transportation Company	1 Transport Drive, Atlanta, GA 23002
CUST002	Finance One Corp	2 Finance Avenue, New York, NY, 10001
CUST003	Cool Book Publishers	3 Book Street, Boston, MA 02110
CUST004	However Forever Energy, Inc.	4 Energy Park, Houston, TX 35004
CUST005	Easy Computing Company	5 Computer Way, Dallas, TX 36543
CUST006	United Suppliers, Inc.	6 Suppliers Street, Chicago, IL 61114
CUST007	Mobile Communications, Inc.	7 Electronics District, Chicago, IL 62643
CUST008	Sports Motor Company	8 Motor Drive, Detroit, MI 55953
CUST009	Swan Stores	9 Riverside Road, Denver, CO 45692
CUST010	Hollywood Studio	10 Media Drive, Los Angeles, CA 78543
CUST011	One Source Technologies, Inc.	11 Technology Way, San Francisco, CA 73285
CUST012	Airspace Industries, Inc.	12 Air Lane, Seattle, WA 83476

TABLE 1.3
SALES
ORGANIZATION

Sales Region	Sales Office	Sales Representative	Sales Representative ID
EAST	ATLANTA	John	SREP01
	NEW YORK	Steve	SREP02
		Mary	SREP03
MIDWEST	DALLAS	Michael	SREP04
		Lisa	SREP05
	CHICAGO	Kevin	SREP06
		Chris	SREP07
WEST	DENVER*	Sam	SREP08
	LOS ANGELES	Eugene	SREP09
	SEATTLE	Mark	SREP10

*Prior to January 1, 2000, the Denver office was in the Midwest region.

You also have three years of sales data, as shown in Table 1.4.

TABLE 1.4 SALES DATA

Customer ID	Sales Representative ID	Material Number	Per Unit Sales Price	Unit of Measure	Quantity Sold	Transaction Date
CUST001	SREP01	MAT001	2	Case	1	19980304
CUST002	SREP02	MAT002	2	Case	2	19990526
CUST002	SREP02	MAT003	5	Case	3	19990730
CUST003	SREP03	MAT003	5	Case	4	20000101
CUST004	SREP04	MAT004	50	Each	5	19991023
CUST004	SREP04	MAT005	100	Each	6	19980904
CUST004	SREP04	MAT005	100	Each	7	19980529
CUST005	SREP05	MAT006	200	Each	8	19991108
CUST006	SREP06	MAT007	20	Each	9	20000408
CUST007	SREP07	MAT008	3	Dozen	10	20000901
CUST007	SREP07	MAT008	3	Dozen	1	19990424
CUST008	SREP08	MAT008	3	Dozen	2	19980328
CUST008	SREP08	MAT009	2	Case	3	19980203
CUST008	SREP08	MAT010	1	U.S. pound	4	19991104
CUST009	SREP09	MAT011	1.5	U.S. pound	5	20000407
CUST010	SREP10	MAT011	1.5	U.S. pound	6	20000701
CUST010	SREP10	MAT011	1.5	U.S. pound	7	19990924
CUST010	SREP10	MAT012	2	U.S. pound	8	19991224
CUST010	SREP10	MAT013	3	Case	9	20000308
CUST011	SREP10	MAT014	1	U.S. pound	10	19980627
CUST012	SREP11	MAT014	2	U.S. pound	1	19991209
CUST012	SREP11	MAT015	3	Case	2	19980221
CUST012	SREP11	MAT015	2	Case	3	20000705
CUST012	SREP11	MAT015	3.5	Case	4	20001225

The data in these tables represent a simplified business scenario. In the real world, you might have years of data and millions of records.

To succeed in the face of fierce market competition, you need to have a complete and up-to-date picture of your business and your business environment. The challenge lies in making the best use of data in decision support. In decision support, you need to perform many kinds of analysis.

This type of **online analytical processing (OLAP)** consumes a lot of computer resources because of the size of data. It cannot be carried out on an **online transaction processing (OLTP)** system, such as a sales management system. Instead, we need a dedicated system, which is the data warehouse.

1.2 Basic Concept of Data Warehousing

A data warehouse is a system with its own database. It draws data from diverse sources and is designed to support query and analysis. To facilitate data retrieval for analytical processing, we use a special database design technique called a star schema.
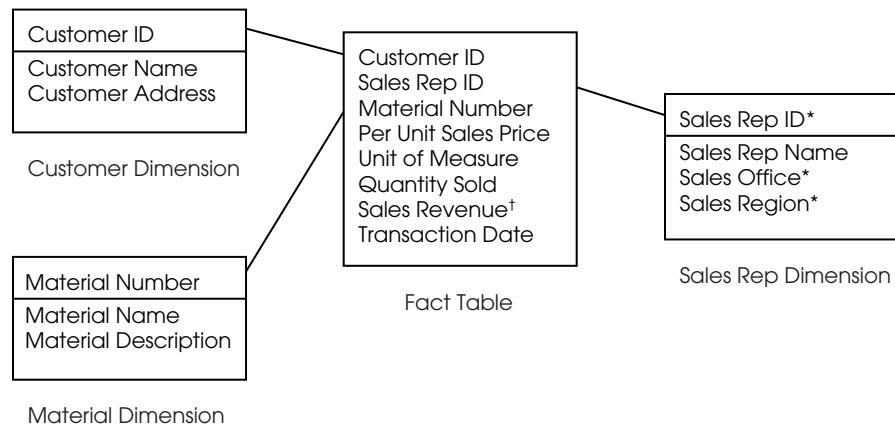
1.2.1 Star Schema

The concept of a star schema is not new; indeed, it has been used in industry for years. For the data in the previous section, we can create a star schema like that shown in Figure 1.1.

The star schema derives its name from its graphical representation—that is, it looks like a star. A **fact table** appears in the middle of the graphic, along with several surrounding **dimension tables**. The central fact table is usually very large, measured in gigabytes. It is the table from which we retrieve the interesting data. The size of the dimension tables amounts to only 1 to 5 percent of the size of the fact table. Common dimensions are unit and time, which are not shown in Figure 1.1. Foreign keys tie the fact table to the dimension tables. Keep in mind that dimension tables are not required to be normalized and that they can contain redundant data.

As indicated in Table 1.3, the sales organization changes over time. The dimension to which it belongs—sales rep dimension—is called the **slowly changing dimension**.

FIGURE 1.1
STAR SCHEMA

Customer ID

Customer Name
Customer Address

Customer Dimension

Customer ID
Sales Rep ID
Material Number
Per Unit Sales Price
Unit of Measure
Quantity Sold
Sales Revenue†
Transaction Date

Fact Table

Sales Rep ID*

Sales Rep Name
Sales Office*
Sales Region*

Sales Rep Dimension

Material Number

Material Name
Material Description

Material Dimension

*Sales Region, Sales Office, and Sales Rep ID are in a hierarchy as shown in Table 1.3.
†Sales Revenue = Per Unit Sales Price $\times$ Quantity Sold.

The following steps explain how a star schema works to calculate the total quantity sold in the Midwest region:

1. From the sales rep dimension, select all sales rep IDs in the Midwest region.
2. From the fact table, select and summarize all quantity sold by the sales rep IDs of Step 1.

### 1.2.2 ETTL—Extracting, Transferring, Transforming, and Loading Data

Besides the difference in designing the database, building a data warehouse involves a critical task that does not arise in building an OLTP system: to extract, transfer, transform, and load (ETTL) data from diverse data sources into the data warehouse (Figure 1.2).

In data extraction, we move data out of source systems, such as an SAP R/3 system. The challenge during this step is to identify the right data. A good knowledge of the source systems is absolutely necessary to accomplish this task.

In data transfer, we move a large amount of data regularly from different source systems to the data warehouse. Here the challenges are to plan a realistic schedule and to have reliable and fast networks.

In data transformation, we format data so that it can be represented consistently in the data warehouse. For example, we might need to convert an entity with multiple names (such as AT&T, ATT, or Bell) into an entity with a single

FIGURE 1.2
ETTL PROCESS
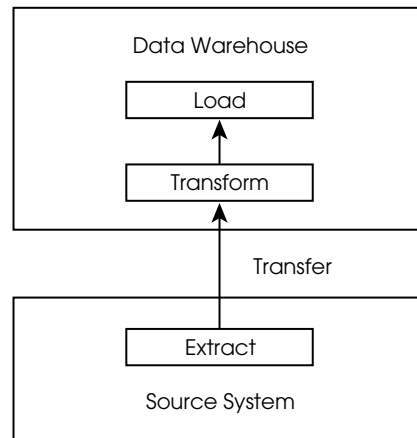
name (such as AT&T). The original data might reside in different databases using different data types, or in different file formats in different file systems. Some are case sensitive; others may be case insensitive.

In data loading, we load data into the fact tables correctly and quickly. The challenge at this step is to develop a robust error-handling procedure.

ETTL is a complex and time-consuming task. Any error can jeopardize data quality, which directly affects business decision making. Because of this fact and for other reasons, most data warehousing projects experience difficulties finishing on time or on budget.

To get a feeling for the challenges involved in ETTL, let's study SAP R/3 as an example. SAP R/3 is a leading ERP (Enterprise Resources Planning) system. According to SAP, the SAP R/3 developer, as of October 2000, some 30,000 SAP R/3 systems were installed worldwide that had 10 million users. SAP R/3 includes several modules, such as SD (sales and distribution), MM (materials management), PP (production planning), FI (financial accounting), and HR (human resources). Basically, you can use SAP R/3 to run your entire business.

SAP R/3's rich business functionality leads to a complex database design. In fact, this system has approximately 10,000 database tables. In addition to the complexity of the relations among these tables, the tables and their columns sometimes don't even have explicit English descriptions. For many years, using the SAP R/3 data for business decision support had been a constant problem.

Recognizing this problem, SAP decided to develop a data warehousing solution to help its customers. The result is SAP Business Information Warehouse, or BW. Since the announcement of its launch in June 1997, BW has drawn intense interest. According to SAP, as of October 2000, more than 1000 SAP BW systems were installed worldwide.

In this book, we will demonstrate how SAP BW implements the star schema and tackles the ETTL challenges.

1.3 BW—An SAP Data Warehousing Solution

BW is an end-to-end data warehousing solution that uses preexisting SAP technologies. BW is built on the Basis 3-tier architecture and coded in the ABAP (Advanced Business Application Programming) language. It uses ALE (Application Link Enabling) and BAPI (Business Application Programming Interface) to link BW with SAP systems and non-SAP systems.

1.3.1 BW Architecture

Figure 1.3 shows the BW architecture at the highest level. This architecture has three layers:

1. The top layer is the reporting environment. It can be **BW Business Explorer (BEx)** or a third-party reporting tool. BEx consists of two components:
    • **BEx Analyzer**
    • **BEx Browser**
    BEx Analyzer is Microsoft Excel with a BW add-in. Thanks to its easy-to-use graphical interface, it allows users to create queries without coding SQL statements. BEx Browser works much like an information center, allowing users to organize and access all kinds of information. Third-party reporting tools connect with BW OLAP Processor through **ODBO (OLE DB for OLAP)**.

FIGURE 1.3
BW
ARCHITECTURE
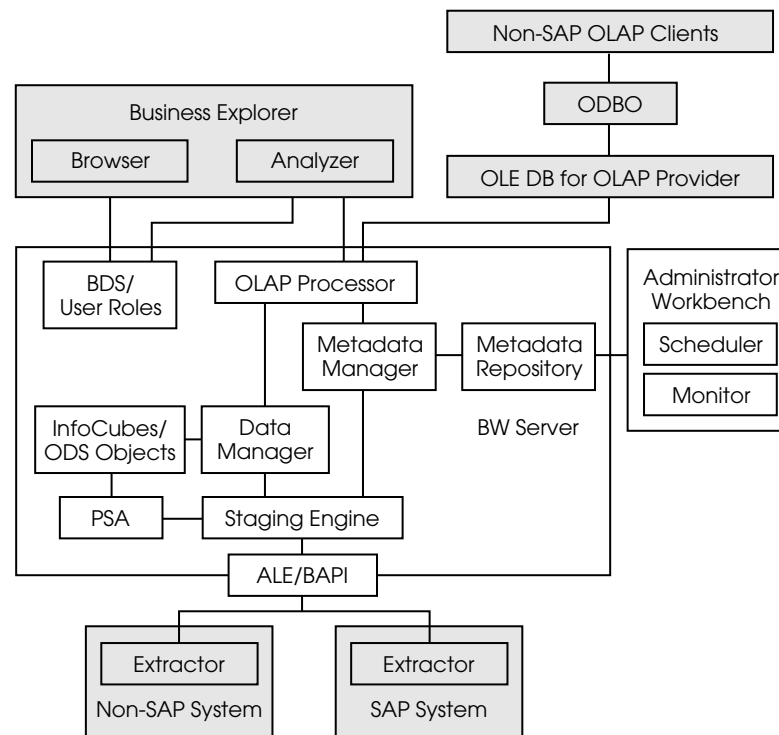
Source: Adapted from SAP BW online documentation.

2. The middle layer, **BW Server**, carries out three tasks:
   - Administering the BW system
   - Storing data
   - Retrieving data according to users' requests
   We will detail BW Server's components next.
3. The bottom layer consists of **source systems**, which can be R/3 systems, BW systems, flat files, and other systems. If the source systems are SAP systems, an SAP component called **Plug-In** must be installed in the source systems. The Plug-In contains **extractors**. An extractor is a set of ABAP programs, database tables, and other objects that BW uses to extract data from the SAP systems. BW connects with SAP systems (R/3 or BW) and flat files via ALE; it connects with non-SAP systems via BAPI.
   The middle-layer BW Server consists of the following components:
   - Administrator Workbench, including BW Scheduler and BW Monitor
   - Metadata Repository and Metadata Manager
   - Staging Engine
   - PSA (Persistent Staging Area)
   - ODS (Operational Data Store) Objects
   - InfoCubes
   - Data Manager
   - OLAP Processor
   - BDS (Business Document Services)
   - User Roles

Administrator Workbench maintains meta-data and all BW objects. It has two components:

- **BW Scheduler** for scheduling jobs to load data
- **BW Monitor** for monitoring the status of data loads

This book mainly focuses on Administrator Workbench.

Metadata Repository contains information about the data warehouse. Meta-data comprise data about data. Metadata Repository contains two types of meta-data: business-related (for example, definitions and descriptions used for reporting) and technical (for example, structure and mapping rules used for data extraction and transformation). We use **Metadata Manager** to maintain Metadata Repository.

Staging Engine implements data mapping and transformation. Triggered by BW Scheduler, it sends requests to a source system for data loading. The source system then selects and transfers data into BW.

PSA (Persistent Staging Area) stores data in the original format while being imported from the source system. PSA allows for quality check before the data are loaded into their destinations, such as ODS Objects or InfoCubes.

ODS (Operational Data Store) Objects allow us to build a multilayer structure for operational data reporting. They are not based on the star schema and are used primarily for detail reporting, rather than for dimensional analysis.

InfoCubes are the fact tables and their associated dimension tables in a star schema.

Data Manager maintains data in ODS Objects and InfoCubes and tells the OLAP Processor what data are available for reporting.

OLAP Processor is the analytical processing engine. It retrieves data from the database, and it analyzes and presents those data according to users' requests.

BDS (Business Document Services) stores documents. The documents can appear in various formats, such as Microsoft Word, Excel, PowerPoint, PDF, and HTML. BEx Analyzer saves query results, or MS Excel files, as workbooks in the BDS.

User Roles are a concept used in SAP authorization management. BW organizes BDS documents according to User Roles. Only users assigned to a particular User Role can access the documents associated with that User Role.

Table 1.5 indicates where each of these components is discussed in this book. As noted in the Preface, this book does *not* discuss third-party reporting tools and BAPI.

TABLE 1.5
CHAPTERS
DETAILING BW
COMPONENTS

Components	Chapters
Business Explorer:	Chapter 5, Creating Queries and Workbooks
• Analyzer and Browser	
Non-SAP OLAP Clients	Not covered
ODBO	
OLE DB for OLAP Provider	
Extractor:	Chapter 3, Loading Data into the InfoCube, on how
• ALE	to load data from flat files
	Chapter 10, Business Content, on how to load data from R/3 systems
	Chapter 11, Generic R/3 Data Extraction
BAPI	Not covered
Administrator Workbench	The entire book, although not explicitly mentioned
BW Scheduler	Chapter 3, Loading Data into the InfoCube, on BW Scheduler
BW Monitor	Chapter 4, Checking Data Quality, on BW Monitor

Components	Chapters
Metadata Repository	The entire book, although not explicitly mentioned
Metadata Manager	
Staging Engine	Chapter 3, Loading Data into the InfoCube
PSA	
	Chapter 4, Checking Data Quality
ODS Objects	Chapter 9, Operational Data Store (ODS)
InfoCubes	Chapter 2, Creating an InfoCube
	Chapter 7, InfoCube Design
	Chapter 8, Aggregates and Multi-Cubes
Data Manager	Chapter 12, Data Maintenance
OLAP Processor	Chapter 13, Performance Tuning
BDS	Chapter 5, Creating Queries and Workbooks
User Roles	Chapter 6, Managing User Authorization

1.3.2 BW Business Content

One of the BW's strongest selling points is its **Business Content**. Business Content contains standard reports and other associated objects. For example, BW provides you, the sales manager, with the following standard reports:

Quotation Processing
- Quotation success rates per sales area
- Quotation tracking per sales area
- General quotation information per sales area

Order Processing
- Monthly incoming orders and revenue
- Sales values
- Billing documents
- Order, delivery, and sales quantities
- Fulfillment rates
- Credit memos
- Proportion of returns to incoming orders
- Returns per customer
- Quantity and values of returns
- Product analysis
- Product profitability analysis

Delivery
- Delivery delays per sales area
- Average delivery processing times

Analyses and Comparisons
- Sales/cost analysis
- Top customers
- Distribution channel analysis
- Product profitability analysis
- Weekly deliveries
- Monthly deliveries
- Incoming orders analysis
- Sales figures comparison
- Returns per customer
- Product analysis
- Monthly incoming orders and revenue

Administrative and Management Functions
- Cost center: plan/actual/variance
- Cost center: responsible for orders, projects, and networks
- Order reports
- WBS Element: plan/actual/variance
- Cost center: plan/actual/variance
- Cost center: hit list of actual variances
- Cost center: actual costs per quarter
- Cost center: capacity-related headcount

Chapter 10 discusses Business Content in detail.

When necessary, we can also use a function, called **Generic Data Extraction**, to extract R/3 data that cannot be extracted with the standard Business Content. Chapter 11 discusses this function in detail.

1.3.3 BW in mySAP.com

BW is evolving rapidly. Knowing its future helps us plan BW projects and their scopes. Here, we give a brief overview of BW's position in mySAP.com.

mySAP.com is SAP's e-business platform that aims to achieve the collaboration among businesses using the Internet technology. It consists of three components:
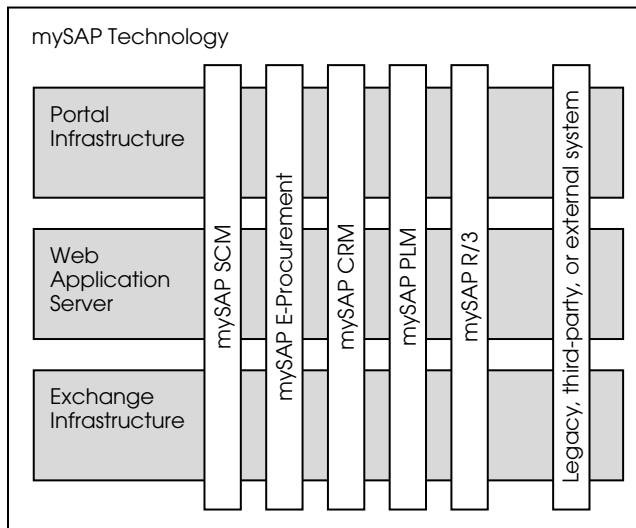
- mySAP Technology
- mySAP Services
- mySAP Hosted Solutions

As shown in Figure 1.4, mySAP Technology includes a portal infrastructure for user-centric collaboration, a Web Application Server for providing Web services, and an exchange infrastructure for process-centric collaboration. The portal infrastructure has a component called mySAP Business Intelligence; it is the same BW but is located in the mySAP.com platform. Using mySAP Technology, SAP develops e-business solutions, such as mySAP Supply Chain Management (mySAP SCM), mySAP Customer Relationship Management (mySAP CRM), and mySAP Product Lifecycle Management (mySAP PLM).

mySAP Services are the services and support that SAP offers to its customers. They range from business analysis, technology implementation, and training to system support. The services and support available from http://service.sap.com/bw/ are good examples of mySAP Services.

mySAP Hosted Solutions are the outsourcing services from SAP. With these solutions, customers do not need to maintain physical machines and networks.

FIGURE 1.4
MYSAP
TECHNOLOGY
AND MYSAP
SOLUTIONS

Source: Adapted from SAP white paper, "mySAP Technology for Open E-Business Integration—Overview."

1.4 Summary

This chapter introduced the basic concept of data warehousing and discussed what SAP BW is, why we need it, its architecture, and what Business Content is. Later chapters will discuss these subjects in more details.

Key Terms

Term	Description
Data warehouse	A data warehouse is a dedicated reporting and analysis environment based on the star schema database design technique that requires paying special attention to the data ETTL process.
Star schema	A star schema is a technique used in the data warehouse database design that aims to help data retrieval for online analytical processing.
ETTL	ETTL represents one of the most challenging tasks in building a data warehouse. It involves the process of extracting, transforming, transferring, and loading data correctly and quickly.
BW	BW is a data warehousing solution from SAP.

For Further Information

- Book: *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses,* by Ralph Kimball. John Wiley & Sons, 1998. ISBN: 0471255475.
- Book: *Business Information Warehouse for SAP,* by Naeem Hashmi. Prima Publishing, 2000. ISBN: 0761523359.
- Web site: http://service.sap.com/bw/ is the official BW Web site maintained by SAP. It contains the original BW materials.
- SAP white paper: "mySAP Technology for Open E-Business Integration—Overview." Available at http://www.sap.com/.
- SAP white paper: "Portal Infrastructure: People-Centric Collaboration." Available at http://www.sap.com/.
- SAP white paper: "mySAP Business Intelligence." Available at http://www.sap.com/.

Next . . .

We will create an InfoCube that implements the Figure 1.1 star schema.