# MODERN DATABASE MANAGEMENT

SIXTH EDITION

JEFFREY A. HOFFER
*University of Dayton*

MARY B. PRESCOTT
*University of Tampa*

FRED R. McFADDEN
*University of Colorado—Colorado Springs*

Prentice Hall

Upper Saddle River, New Jersey 07458

Executive Editor: Bob Horan
Senior Executive Editor: David Alexander
Publisher: Natalie Anderson
Project Manager: Lori Cerreto
Editorial Assistant: Maat Van Uitert
Media Project Manager: Joan Waxman
Senior Marketing Manager: Sharon Turkovich
Marketing Assistant: Scott Patterson
Managing Editor (Production): Cynthia Regan
Production Editor: Michael Reynolds
Production Assistant: Dianne Falcone
Permissions Supervisor: Suzanne Grappi
Associate Director, Manufacturing: Vincent Scelta
Production Manager: Arnold Vila
Design Director: Patricia Smythe
Interior Design: Lee Goldstein
Cover Design: Steven Frim
Cover Painting: Steven Frim
Illustrator (Interior): Electragraphics
Associate Director, Multimedia Production: Karen Goldsmith
Manager, Multimedia Production: Christy Mahon
Print Production Liaison: Ashley Scattergood
Composition: UG / GGS Information Services, Inc.
Full-Service Project Management: UG / GGS Information Services, Inc.
Printer/Binder: R. R. Donnelley, Willard

**Library of Congress Cataloging-in-Publication Data**

Hoffer, Jeffrey A.,
Modern database management / Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden.—6th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-033969-5
1. Database management. I. Hoffer, Jeffrey A. II. Prescott, Mary B.
III. Title
QA76.9.D3 M395 2001
005.74—dc21 2001036826

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on page 626.



Prentice Hall

10 9 8 7 6 5 4 3 2
ISBN 0-13-033969-5

building separate databases, called "data warehouses," for this type of decision support application (Lambert, 1996).

Use of databases to support customer relationship management, on-line shopping, and employee relationship management is increasingly important. Databases are fundamental to most information systems now, from small databases used in personal digital assistants and information appliances to the very large databases that support enterprise-wide information systems.

Although the future of databases is assured, much work remains to be done. Many organizations have a proliferation of incompatible databases that were developed to meet immediate needs, rather than based on a planned strategy or a well-managed evolution. Much of the data are trapped in older, "legacy" systems, and the data are often of poor quality. New skills are required to design data warehouses, and there is a critical shortage of skills in areas such as database analysis, database design, data administration, and database administration. We address these and other important issues in this textbook.

A course in database management has emerged as one of the most important courses in the information systems curriculum today. As an information systems professional, you must be prepared to analyze database requirements and design and implement databases within the context of information systems development. You must be prepared as well to consult with end-users and show them how they can use databases (or data warehouses) to build decision support systems and executive information systems for competitive advantage. And, the widespread use of databases attached to Websites in order to return dynamic information to users of the Website requires that you understand not only how to attach databases to the Web, but also how to secure those databases so that their contents may be viewed but not compromised by outside users.

In this chapter we introduce the basic concepts of databases and database management systems (DBMS). We describe traditional file management systems and some of their shortcomings that led to the database approach. We describe the range of database applications, from personal computers and digital assistants to workgroup, departmental, and enterprise databases. Next we consider the benefits, costs, and risks of using the database approach. We conclude the chapter with a summary of the evolution of database systems and of the range of technologies used to build, use, and manage databases. This chapter is intended to serve as a preview of the topics in the remainder of the text.

## BASIC CONCEPTS AND DEFINITIONS

**Database:** An organized collection of logically related data.

We define a **database** as an organized collection of logically related data. A database may be of any size and complexity. For example, a salesperson may maintain a small database of customer contacts on her laptop computer that consists of a few megabytes of data. A large corporation may build a very large database consisting of several terabytes of data (a *terabyte* is a trillion bytes) on a large mainframe computer that is used for decision support applications (Winter, 1997). Very large data warehouses contain more than a petabyte of data (a *petabyte* is a quadrillion bytes). (We assume throughout the text that all databases are computer-based.)

### Data

Historically, the term *data* referred to known facts that could be recorded and stored on computer media. For example in a salesperson's database, the data would include facts such as customer name, address, and telephone number. This defini-

### Conversion Costs

The term *legacy systems* is widely used to refer to older applications in an org zation that are based on file processing and/or older database technology. The of converting these older systems to modern database technology—measure terms of dollars, time, and organizational commitment—may often seem prohib to an organization. As will be shown in Chapter 11, the use of data warehouses is strategy for continuing to use older systems while at the same time exploiting m ern database technology and techniques (Ritter, 1999).

### Need for Explicit Backup and Recovery

A shared corporate database must be accurate and available at all times. T requires that comprehensive procedures be developed and used for provid backup copies of data and for restoring a database when damage occurs. A mod database management system normally automates many more of the backup recovery tasks than a file system. We describe procedures for security, backup, recovery in Chapter 12.

### Organizational Conflict

A shared database requires a consensus on data definitions and ownershi well as responsibilities for accurate data maintenance. Experience has shown t conflicts on data definitions, data formats and coding, rights to update shared d and associated issues are frequent and often difficult to resolve. Handling th issues requires organizational commitment to the database approach, organizati ally astute database administrators, and a sound evolutionary approach to datab development.

If strong top management support of and commitment to the database appro is lacking, end-user development of stand-alone databases is likely to prolifera These databases do not follow the general database approach that we have describ and they are unlikely to provide the benefits described earlier.

## COMPONENTS OF THE DATABASE ENVIRONMENT

The major components of a typical database environment and their relati ships are shown in Figure 1-10. You have already been introduced to some (but n all) of these components in previous sections. Following is a brief description of t nine components shown in Figure 1-10.

1. **Computer-aided software engineering (CASE) tools** Automated tools us to design databases and application programs. We describe the use of CA tools for database design and development throughout the text.
2. **Repository** Centralized knowledge base for all data definitions, data re tionships, screen and report formats, and other system components. A rep itory contains an extended set of metadata important for managing databa as well as other components of an information system. We describe the rep itory in Chapter 12.
3. **Database management system (DBMS)** Commercial software (and occasio ally, hardware and firmware) system used to define, create, maintain, a provide controlled access to the database and also to the repository. W describe the functions of a DBMS in Chapters 12 and 13.

**Repository:** A centralized knowledge base of all data definitions, data relationships, screen and report formats, and other system components.

**Database management system (DBMS):** A software application that is used to create, maintain, and provide controlled access to user databases.

tional data model.) We next describe and illustrate the process of transforming E-R model into the relational model. Many CASE tools support this transform today; however, it is important that you understand the underlying principles procedures. We then describe the concepts of normalization in detail. Normali tion, which is the process of designing well-structured relations, is an important ponent of logical design for the relational model. Finally, we describe how to m relations while avoiding common pitfalls that may occur in this process.

The objective of logical database design is to translate the conceptual des (which represents an organization's requirements for data) into a logical datab design that can be implemented on a chosen database management system. Th resulting databases must meet user needs for data sharing, flexibility, and ease access. The concepts presented in this chapter are essential to your understanding the database development process.

## THE RELATIONAL DATA MODEL

The relational data model was first introduced in 1970 by E. F. Codd, then IBM (Codd, 1970). Two early research projects were launched to prove the feasibil of the relational model and to develop prototype systems. The first of these, at IBM San Jose Research Laboratory, led to the development of System R (a prototype re tional DBMS–RDBMS) during the late 1970s. The second, at the University of Ca fornia at Berkeley, led to the development of Ingres, an academically orien RDBMS. Commercial RDBMS products from numerous vendors started to appe about 1980 (see the Website for this book for links to RDBMS and other DBMS ve dors). Today RDBMSs have become the dominant technology for database manag ment, and there are literally hundreds of RDBMS products for computers rang from personal computers to mainframes.

### Basic Definitions

The relational data model represents data in the form of tables. The relation model is based on mathematical theory and therefore has a solid theoretical founda tion. However, we need only a few simple concepts to describe the relational mode and it is therefore easily understood and used by those unfamiliar with the underl ing theory. The relational data model consists of the following three componen (Fleming and von Halle, 1989):

1. *Data structure* Data are organized in the form of tables with rows an columns.
2. *Data manipulation* Powerful operations (using the SQL language) are use to manipulate data stored in the relations.
3. *Data integrity* Facilities are included to specify business rules that maintain the integrity of data when they are manipulated.

We discuss data structure and data integrity in this section. Data manipulation is dis cussed in Chapters 7, 8, and 10.

**Relation:** A named two-dimensional table of data.

**Relational Data Structure** A **relation** is a named, two-dimensional table of data. Each relation (or table) consists of a set of named columns and an arbitrary number of unnamed rows. An attribute, consistent with its definition in Chapter 3, is a named column of a relation. Each row of a relation corresponds to a record that contains data (attribute) values for a single entity. Figure 5-1 shows an example of a relation named EMPLOYEE1. This relation contains the following attributes describing employees: Emp_ID, Name, Dept_Name, and Salary. The five rows of the table corre spond to five employees. It is important to understand that the sample data in Figure