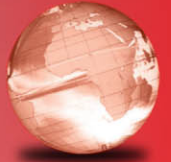


GLOBAL
EDITION



Modern Database Management

TWELFTH EDITION

Jeff Hoffer • Ramesh Venkataraman • Heikki Topi

ALWAYS LEARNING

PEARSON

Global Edition

Twelfth Edition

MODERN DATABASE MANAGEMENT

This page intentionally left blank

Global Edition

Twelfth Edition

MODERN DATABASE MANAGEMENT

Jeffrey A. Hoffer

University of Dayton

V. Ramesh

Indiana University

Heikki Topi

Bentley University

PEARSON

Boston Columbus Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Vice President, Business Publishing: Donna Battista
Editor-in-Chief: Stephanie Wall
Acquisitions Editor: Nicole Sam
Program Manager Team Lead: Ashley Santora
Program Manager: Denise Weiss
Editorial Assistant: Olivia Vignone
Vice President, Product Marketing: Maggie Moylan
Director of Marketing, Digital Services and Products:
Jeanette Koskinas
Field Marketing Manager: Lenny Ann Raper
Senior Strategic Marketing Manager: Erin Gardner
Product Marketing Assistant: Jessica Quazza
Project Manager Team Lead: Jeff Holcomb
Project Manager: Ilene Kahn
Assistant Acquisitions Editor, Global Edition: Ananya Srivastava
Associate Project Editor, Global Edition: Amrita Kar
Project Manager, Global Edition: Nikhil Rakshit
Manager, Media Production, Global Edition: Vikram Kumar
Senior Manufacturing Controller, Production, Global Edition:
Trudy Kimber

Operations Specialist: Diane Peirano
Creative Director: Blair Brown
Senior Art Director: Janet Slowik
Interior and Cover Designer: Shibu Velayudhan,
Lumina Datamatics Ltd
Cover Image: Neale Cousland/Shutterstock
Vice President, Director of Digital Strategy & Assessment:
Paul Gentile
Manager of Learning Applications: Paul Deluca
Digital Editor: Brian Surette
Digital Studio Manager: Diane Lombardo
Digital Studio Project Manager: Robin Lazrus
Digital Studio Project Manager: Alana Coles
Digital Studio Project Manager: Monique Lawrence
Digital Studio Project Manager: Regina DaSilva
Full-Service Project Management and Composition:
George Jacob/Integra Software Solutions Pvt., Ltd.
Printer/Binder: Vivar
Cover Printer: Vivar
Text Font: 10/12 PalatinoLTStd Roman

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within text.

Microsoft and/or its respective suppliers make no representations about the suitability of the information contained in the documents and related graphics published as part of the services for any purpose. All such documents and related graphics are provided “as is” without warranty of any kind. Microsoft and/or its respective suppliers hereby disclaim all warranties and conditions with regard to this information, including all warranties and conditions of merchantability, whether express, implied or statutory, fitness for a particular purpose, title and non-infringement. In no event shall Microsoft and/or its respective suppliers be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of information available from the services.

The documents and related graphics contained herein could include technical inaccuracies or typographical errors. Changes are periodically added to the information herein. Microsoft and/or its respective suppliers may make improvements and/or changes in the product(s) and/or the program(s) described herein at any time. Partial screen shots may be viewed in full within the software version specified.

Trademarks

Microsoft® Windows®, and Microsoft Office® are registered trademarks of the Microsoft Corporation in the U.S.A. and other countries. This book is not sponsored or endorsed by or affiliated with the Microsoft Corporation.

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com.

© Pearson Education Limited 2016

The rights of Jeff Hoffer, Ramesh Venkataraman and Heikki Topi to be identified as the authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Modern Database Management, 12th edition, ISBN 978-0-13-354461-9, by Jeff Hoffer, Ramesh Venkataraman and Heikki Topi, published by Pearson Education © 2016.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC 1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

ISBN-10: 1-292-10185-7
ISBN-13: 978-1-292-10185-9

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

10 9 8 7 6 5 4 3 2 1

Typeset by Integra Software Solutions Pvt., Ltd.
Printed and bound by Vivar

To Patty, for her sacrifices, encouragement, and support for more than 30 years of being a textbook author widow. To my students and colleagues, for being receptive and critical and for challenging me to be a better teacher.

—J.A.H.

To Gayathri, for her sacrifices and patience these past 25 years. To my parents, for letting me make the journey abroad, and to my cat, Raju, who was a part of our family for more than 20 years.

—V.R.

To Anne-Louise, for her loving support, encouragement, and patience. To Leila and Saara, whose laughter and joy of life continue to teach me about what is truly important. To my teachers, colleagues, and students, from whom I continue to learn every day.

—H.T.

This page intentionally left blank

BRIEF CONTENTS

Part I The Context of Database Management 37

Chapter 1 The Database Environment and Development Process 38

Part II Database Analysis 87

Chapter 2 Modeling Data in the Organization 89

Chapter 3 The Enhanced E-R Model 150

Part III Database Design 189

Chapter 4 Logical Database Design and the Relational Model 191

Chapter 5 Physical Database Design and Performance 242

Part IV Implementation 277

Chapter 6 Introduction to SQL 279

Chapter 7 Advanced SQL 325

Chapter 8 Database Application Development 373

Chapter 9 Data Warehousing 410

Part V Advanced Database Topics 453

Chapter 10 Data Quality and Integration 455

Chapter 11 Big Data and Analytics 481

Chapter 12 Data and Database Administration 521

Glossary of Acronyms 570

Glossary of Terms 572

Index 580

Available Online at www.pearsonhighered.com/hoffer

Chapter 13 Distributed Databases 13-1

Chapter 14 Object-Oriented Data Modeling 14-1

Appendices

Appendix A Data Modeling Tools and Notation A-37

Appendix B Advanced Normal Forms B-37

Appendix C Data Structures C-37

This page intentionally left blank

CONTENTS

Preface 25

Part I The Context of Database Management 37

An Overview of Part One 37

Chapter 1 The Database Environment and Development Process 38

Learning Objectives 38

Data Matter! 38

Introduction 39

Basic Concepts and Definitions 41

Data 41

Data Versus Information 41

Metadata 42

Traditional File Processing Systems 43

File Processing Systems at Pine Valley Furniture Company 44

Disadvantages of File Processing Systems 44

PROGRAM-DATA DEPENDENCE 44

DUPLICATION OF DATA 45

LIMITED DATA SHARING 45

LENGTHY DEVELOPMENT TIMES 45

EXCESSIVE PROGRAM MAINTENANCE 45

The Database Approach 45

Data Models 45

ENTITIES 46

RELATIONSHIPS 47

Relational Databases 47

Database Management Systems 47

Advantages of the Database Approach 47

PROGRAM-DATA INDEPENDENCE 47

PLANNED DATA REDUNDANCY 48

IMPROVED DATA CONSISTENCY 48

IMPROVED DATA SHARING 48

INCREASED PRODUCTIVITY OF APPLICATION DEVELOPMENT 49

ENFORCEMENT OF STANDARDS 49

IMPROVED DATA QUALITY 49

IMPROVED DATA ACCESSIBILITY AND RESPONSIVENESS 50

REDUCED PROGRAM MAINTENANCE 50

IMPROVED DECISION SUPPORT 50

Cautions About Database Benefits 50

Costs and Risks of the Database Approach 50

NEW, SPECIALIZED PERSONNEL 51

INSTALLATION AND MANAGEMENT COST AND COMPLEXITY 51

CONVERSION COSTS 51

NEED FOR EXPLICIT BACKUP AND RECOVERY 51

ORGANIZATIONAL CONFLICT 51

Components of the Database Environment 51



The Database Development Process 53

- Systems Development Life Cycle 54
 - PLANNING—ENTERPRISE MODELING 54
 - PLANNING—CONCEPTUAL DATA MODELING 54
 - ANALYSIS—CONCEPTUAL DATA MODELING 54
 - DESIGN—LOGICAL DATABASE DESIGN 55
 - DESIGN—PHYSICAL DATABASE DESIGN AND DEFINITION 56
 - IMPLEMENTATION—DATABASE IMPLEMENTATION 56
 - MAINTENANCE—DATABASE MAINTENANCE 56
- Alternative Information Systems (IS) Development Approaches 57
- Three-Schema Architecture for Database Development 58
- Managing the People Involved in Database Development 60

Evolution of Database Systems 60

- 1960s 62
- 1970s 62
- 1980s 62
- 1990s 62
- 2000 and Beyond 63

The Range of Database Applications 63

- Personal Databases 64
- Multitier Client/Server Databases 64
- Enterprise Applications 65

Developing a Database Application for Pine Valley Furniture Company 67

- Database Evolution at Pine Valley Furniture Company 68
- Project Planning 69
- Analyzing Database Requirements 70
- Designing the Database 72
- Using the Database 75
- Administering the Database 76
- Future of Databases at Pine Valley 77
 - Summary 77 • Key Terms 78 • Review Questions 78 • Problems and Exercises 80 • Field Exercises 81 • References 82 • Further Reading 82 • Web Resources 83*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 84**



Part II Database Analysis 87

An Overview of Part Two 87

Chapter 2 Modeling Data in the Organization 89

- Learning Objectives 89
- Introduction 89
- The E-R Model: An Overview 92
 - Sample E-R Diagram 92
 - E-R Model Notation 94
- Modeling the Rules of the Organization 95
 - Overview of Business Rules 96
 - THE BUSINESS RULES PARADIGM 96



- Scope of Business Rules 97
 - GOOD BUSINESS RULES 97
 - GATHERING BUSINESS RULES 98
- Data Names and Definitions 98
 - DATA NAMES 98
 - DATA DEFINITIONS 99
 - GOOD DATA DEFINITIONS 99
- Modeling Entities and Attributes 101
 - Entities 101
 - ENTITY TYPE VERSUS ENTITY INSTANCE 101
 - ENTITY TYPE VERSUS SYSTEM INPUT, OUTPUT, OR USER 101
 - STRONG VERSUS WEAK ENTITY TYPES 102
 - NAMING AND DEFINING ENTITY TYPES 103
 - Attributes 105
 - REQUIRED VERSUS OPTIONAL ATTRIBUTES 105
 - SIMPLE VERSUS COMPOSITE ATTRIBUTES 106
 - SINGLE-VALUED VERSUS MULTIVALUED ATTRIBUTES 106
 - STORED VERSUS DERIVED ATTRIBUTES 107
 - IDENTIFIER ATTRIBUTE 107
 - NAMING AND DEFINING ATTRIBUTES 108
- Modeling Relationships 110
 - Basic Concepts and Definitions in Relationships 111
 - ATTRIBUTES ON RELATIONSHIPS 112
 - ASSOCIATIVE ENTITIES 112
 - Degree of a Relationship 114
 - UNARY RELATIONSHIP 114
 - BINARY RELATIONSHIP 116
 - TERNARY RELATIONSHIP 117
 - Attributes or Entity? 118
 - Cardinality Constraints 120
 - MINIMUM CARDINALITY 120
 - MAXIMUM CARDINALITY 120
 - Some Examples of Relationships and Their Cardinalities 121
 - A TERNARY RELATIONSHIP 122
 - Modeling Time-Dependent Data 122
 - Modeling Multiple Relationships Between Entity Types 125
 - Naming and Defining Relationships 126
- E-R Modeling Example: Pine Valley Furniture Company 128
- Database Processing at Pine Valley Furniture 130
 - Showing Product Information 131
 - Showing Product Line Information 131
 - Showing Customer Order Status 132
 - Showing Product Sales 133
 - Summary 134 • Key Terms 135 • Review Questions 135 • Problems and Exercises 136 • Field Exercises 146 • References 146 • Further Reading 147 • Web Resources 147*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 148**



Chapter 3 The Enhanced E-R Model 150

- Learning Objectives 150
- Introduction 150

Representing Supertypes and Subtypes 151

- Basic Concepts and Notation 152
 - AN EXAMPLE OF A SUPERTYPE/SUBTYPE RELATIONSHIP 153
 - ATTRIBUTE INHERITANCE 154
 - WHEN TO USE SUPERTYPE/SUBTYPE RELATIONSHIPS 154
- Representing Specialization and Generalization 155
 - GENERALIZATION 155
 - SPECIALIZATION 156
 - COMBINING SPECIALIZATION AND GENERALIZATION 157
- Specifying Constraints in Supertype/Subtype Relationships 158
 - Specifying Completeness Constraints 158
 - TOTAL SPECIALIZATION RULE 158
 - PARTIAL SPECIALIZATION RULE 158
 - Specifying Disjointness Constraints 159
 - DISJOINT RULE 159
 - OVERLAP RULE 159
 - Defining Subtype Discriminators 160
 - DISJOINT SUBTYPES 160
 - OVERLAPPING SUBTYPES 161
 - Defining Supertype/Subtype Hierarchies 161
 - AN EXAMPLE OF A SUPERTYPE/SUBTYPE HIERARCHY 162
 - SUMMARY OF SUPERTYPE/SUBTYPE HIERARCHIES 163
- EER Modeling Example: Pine Valley Furniture Company 164
- Entity Clustering 167
- Packaged Data Models 170
 - A Revised Data Modeling Process with Packaged Data Models 172
 - Packaged Data Model Examples 174
 - Summary 179 • Key Terms 180 • Review Questions 180 • Problems and Exercises 181 • Field Exercises 184 • References 184 • Further Reading 184 • Web Resources 185*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 186**



Part III Database Design 189

An Overview of Part Three 189

Chapter 4 Logical Database Design and the Relational Model 191



Learning Objectives 191

Introduction 191

The Relational Data Model 192

Basic Definitions 192

RELATIONAL DATA STRUCTURE 193

RELATIONAL KEYS 193

PROPERTIES OF RELATIONS 194

REMOVING MULTIVALUED ATTRIBUTES FROM TABLES 194

Sample Database 194

Integrity Constraints 196

Domain Constraints 196

Entity Integrity 196

Referential Integrity 198

Creating Relational Tables	199
Well-Structured Relations	200
Transforming EER Diagrams into Relations	201
Step 1: Map Regular Entities	202
COMPOSITE ATTRIBUTES	202
MULTIVALUED ATTRIBUTES	203
Step 2: Map Weak Entities	203
WHEN TO CREATE A SURROGATE KEY	205
Step 3: Map Binary Relationships	205
MAP BINARY ONE-TO-MANY RELATIONSHIPS	205
MAP BINARY MANY-TO-MANY RELATIONSHIPS	206
MAP BINARY ONE-TO-ONE RELATIONSHIPS	206
Step 4: Map Associative Entities	207
IDENTIFIER NOT ASSIGNED	208
IDENTIFIER ASSIGNED	208
Step 5: Map Unary Relationships	209
UNARY ONE-TO-MANY RELATIONSHIPS	209
UNARY MANY-TO-MANY RELATIONSHIPS	210
Step 6: Map Ternary (and n -ary) Relationships	211
Step 7: Map Supertype/Subtype Relationships	212
Summary of EER-to-Relational Transformations	214
Introduction to Normalization	214
Steps in Normalization	215
Functional Dependencies and Keys	215
DETERMINANTS	217
CANDIDATE KEYS	217
Normalization Example: Pine Valley Furniture Company	218
Step 0: Represent the View in Tabular Form	218
Step 1: Convert to First Normal Form	219
REMOVE REPEATING GROUPS	219
SELECT THE PRIMARY KEY	219
ANOMALIES IN 1NF	220
Step 2: Convert to Second Normal Form	221
Step 3: Convert to Third Normal Form	222
REMOVING TRANSITIVE DEPENDENCIES	222
Determinants and Normalization	223
Step 4: Further Normalization	223
Merging Relations	224
An Example	224
View Integration Problems	224
SYNONYMS	225
HOMONYMS	225
TRANSITIVE DEPENDENCIES	225
SUPERTYPE/SUBTYPE RELATIONSHIPS	226
A Final Step for Defining Relational Keys	226
<i>Summary</i>	228
<i>Key Terms</i>	230
<i>Review Questions</i>	230
<i>Problems and Exercises</i>	231
<i>Field Exercises</i>	240
<i>References</i>	240
<i>Further Reading</i>	240
<i>Web Resources</i>	240
▶ CASE: Forondo Artist Management Excellence Inc.	241



Chapter 5 Physical Database Design and Performance 242

- Learning Objectives 242
- Introduction 242
- The Physical Database Design Process 243
 - Physical Database Design as a Basis for Regulatory Compliance 244
 - Data Volume and Usage Analysis 245
- Designing Fields 246
 - Choosing Data Types 247
 - CODING TECHNIQUES 248
 - HANDLING MISSING DATA 249
- Denormalizing and Partitioning Data 249
 - Denormalization 249
 - OPPORTUNITIES FOR AND TYPES OF DENORMALIZATION 250
 - DENORMALIZE WITH CAUTION 252
 - Partitioning 253
- Designing Physical Database Files 255
 - File Organizations 257
 - HEAP FILE ORGANIZATION 257
 - SEQUENTIAL FILE ORGANIZATIONS 257
 - INDEXED FILE ORGANIZATIONS 257
 - HASHED FILE ORGANIZATIONS 260
 - Clustering Files 263
 - Designing Controls for Files 263
- Using and Selecting Indexes 264
 - Creating a Unique Key Index 264
 - Creating a Secondary (Nonunique) Key Index 264
 - When to Use Indexes 265
- Designing a Database for Optimal Query Performance 266
 - Parallel Query Processing 266
 - Overriding Automatic Query Optimization 267
 - Summary 268 • Key Terms 269 • Review Questions 269 • Problems and Exercises 270 • Field Exercises 273 • References 273 • Further Reading 273 • Web Resources 274*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 275**

**Part IV Implementation 277**

An Overview of Part Four 277

**Chapter 6 Introduction to SQL 279**

- Learning Objectives 279
- Introduction 279
- Origins of the SQL Standard 281
- The SQL Environment 283
- Defining a Database in SQL 287
 - Generating SQL Database Definitions 288
 - Creating Tables 289
 - Creating Data Integrity Controls 291
 - Changing Table Definitions 292
 - Removing Tables 293

Inserting, Updating, and Deleting Data	293
Batch Input	295
Deleting Database Contents	295
Updating Database Contents	295
Internal Schema Definition in RDBMSs	296
Creating Indexes	296
Processing Single Tables	297
Clauses of the SELECT Statement	298
Using Expressions	300
Using Functions	301
Using Wildcards	303
Using Comparison Operators	303
Using Null Values	304
Using Boolean Operators	304
Using Ranges for Qualification	307
Using Distinct Values	307
Using IN and NOT IN with Lists	309
Sorting Results: The ORDER BY Clause	310
Categorizing Results: The GROUP BY Clause	311
Qualifying Results by Categories: The HAVING Clause	312
Using and Defining Views	313
MATERIALIZED VIEWS	317
<i>Summary</i>	317
<i>Key Terms</i>	318
<i>Review Questions</i>	318
<i>Problems and Exercises</i>	319
<i>Field Exercises</i>	322
<i>References</i>	323
<i>Further Reading</i>	323
<i>Web Resources</i>	323
▶ CASE: Forondo Artist Management Excellence Inc.	324



Chapter 7 Advanced SQL 325

Learning Objectives	325
Introduction	325
Processing Multiple Tables	326
Equi-join	327
Natural Join	328
Outer Join	329
Sample Join Involving Four Tables	331
Self-Join	333
Subqueries	334
Correlated Subqueries	339
Using Derived Tables	341
Combining Queries	342
Conditional Expressions	344
More Complicated SQL Queries	344
Tips for Developing Queries	346
Guidelines for Better Query Design	348
Ensuring Transaction Integrity	350
Data Dictionary Facilities	351
Recent Enhancements and Extensions to SQL	353
Analytical and OLAP Functions	353
New Data Types	355



- New Temporal Features in SQL 355
- Other Enhancements 356
- Triggers and Routines 357
 - Triggers 357
 - Routines and other Programming Extensions 359
 - Example Routine in Oracle's PL/SQL 361
- Embedded SQL and Dynamic SQL 363
 - Summary 365 • Key Terms 366 • Review Questions 366 • Problems and Exercises 367 • Field Exercises 370 • References 370 • Further Reading 370 • Web Resources 371*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 372**

Chapter 8 Database Application Development 373

- Learning Objectives 373
- Location, Location, Location! 373
- Introduction 374
- Client/Server Architectures 374
- Databases in a Two-Tier Architecture 376
 - A VB.NET Example 378
 - A Java Example 380
- Three-Tier Architectures 381
- Web Application Components 383
- Databases in Three-Tier Applications 385
 - A JSP Web Application 385
 - A PHP Example 389
 - An ASP.NET Example 391
- Key Considerations in Three-Tier Applications 392
 - Stored Procedures 392
 - Transactions 395
 - Database Connections 395
 - Key Benefits of Three-Tier Applications 395
 - Cloud Computing and Three-Tier Applications 396
- Extensible Markup Language (XML) 397
 - Storing XML Documents 399
 - Retrieving XML Documents 399
 - Displaying XML Data 402
 - XML and Web Services 402
 - Summary 405 • Key Terms 406 • Review Questions 406 • Problems and Exercises 407 • Field Exercises 408 • References 408 • Further Reading 408 • Web Resources 408*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 409**



Chapter 9 Data Warehousing 410

- Learning Objectives 410
- Introduction 410
- Basic Concepts of Data Warehousing 412
 - A Brief History of Data Warehousing 413
 - The Need for Data Warehousing 413
 - NEED FOR A COMPANY-WIDE VIEW 413
 - NEED TO SEPARATE OPERATIONAL AND INFORMATIONAL SYSTEMS 415

Data Warehouse Architectures	416
Independent Data Mart Data Warehousing Environment	416
Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach	418
Logical Data Mart and Real-Time Data Warehouse Architecture	420
Three-Layer Data Architecture	423
ROLE OF THE ENTERPRISE DATA MODEL	424
ROLE OF METADATA	424
Some Characteristics of Data Warehouse Data	424
Status Versus Event Data	424
Transient Versus Periodic Data	425
An Example of Transient and Periodic Data	425
TRANSIENT DATA	425
PERIODIC DATA	427
OTHER DATA WAREHOUSE CHANGES	427
The Derived Data Layer	428
Characteristics of Derived Data	428
The Star Schema	429
FACT TABLES AND DIMENSION TABLES	429
EXAMPLE STAR SCHEMA	430
SURROGATE KEY	431
GRAIN OF THE FACT TABLE	432
DURATION OF THE DATABASE	433
SIZE OF THE FACT TABLE	433
MODELING DATE AND TIME	434
Variations of the Star Schema	435
MULTIPLE FACT TABLES	435
FACTLESS FACT TABLES	436
Normalizing Dimension Tables	437
MULTIVALUED DIMENSIONS	437
HIERARCHIES	438
Slowly Changing Dimensions	440
Determining Dimensions and Facts	442
The Future of Data Warehousing: Integration with Big Data and Analytics	444
Speed of Processing	445
Cost of Storing Data	445
Dealing with Unstructured Data	445
<i>Summary</i>	446
<i>Key Terms</i>	446
<i>Review Questions</i>	447
<i>Problems and Exercises</i>	447
<i>Field Exercises</i>	451
<i>References</i>	451
<i>Further Reading</i>	452
<i>Web Resources</i>	452

Part V Advanced Database Topics 453

An Overview of Part Five 453

Chapter 10 Data Quality and Integration 455

Learning Objectives 455

Introduction 455

Data Governance 456

Managing Data Quality	457
Characteristics of Quality Data	458
EXTERNAL DATA SOURCES	459
REDUNDANT DATA STORAGE AND INCONSISTENT METADATA	460
DATA ENTRY PROBLEMS	460
LACK OF ORGANIZATIONAL COMMITMENT	460
Data Quality Improvement	460
GET THE BUSINESS BUY-IN	460
CONDUCT A DATA QUALITY AUDIT	461
ESTABLISH A DATA STEWARDSHIP PROGRAM	462
IMPROVE DATA CAPTURE PROCESSES	462
APPLY MODERN DATA MANAGEMENT PRINCIPLES AND TECHNOLOGY	463
APPLY TQM PRINCIPLES AND PRACTICES	463
Summary of Data Quality	463
Master Data Management	464
Data Integration: An Overview	465
General Approaches to Data Integration	465
DATA FEDERATION	466
DATA PROPAGATION	467
Data Integration for Data Warehousing: The Reconciled Data Layer	467
Characteristics of Data After ETL	467
The ETL Process	468
MAPPING AND METADATA MANAGEMENT	468
EXTRACT	469
CLEANSE	470
LOAD AND INDEX	472
Data Transformation	473
Data Transformation Functions	474
RECORD-LEVEL FUNCTIONS	474
FIELD-LEVEL FUNCTIONS	475
Summary	477
Key Terms	477
Review Questions	477
Problems and Exercises	478
Field Exercises	479
References	479
Further Reading	480
Web Resources	480

Chapter 11 Big Data and Analytics 481

Learning Objectives	481
Introduction	481
Big Data	483
NoSQL	485
Classification of NoSQL Database Management Systems	486
KEY-VALUE STORES	486
DOCUMENT STORES	486
WIDE-COLUMN STORES	487
GRAPH-ORIENTED DATABASES	487
NoSQL Examples	488
REDIS	488
MONGODB	488
APACHE CASSANDRA	488
NEO4J	488
Impact of NoSQL on Database Professionals	488

Hadoop	489
Components of Hadoop	490
THE HADOOP DISTRIBUTED FILE SYSTEM (HDFS)	490
MAPREDUCE	491
PIG	492
HIVE	492
HBASE	493
Integrated Analytics and Data Science Platforms	493
HP HAVEN	493
TERADATA ASTER	493
IBM BIG DATA PLATFORM	493
Putting It All Together: Integrated Data Architecture	494
Analytics	496
Types of Analytics	497
Use of Descriptive Analytics	498
SQL OLAP QUERYING	499
ONLINE ANALYTICAL PROCESSING (OLAP) TOOLS	501
DATA VISUALIZATION	503
BUSINESS PERFORMANCE MANAGEMENT AND DASHBOARDS	505
Use of Predictive Analytics	506
DATA MINING TOOLS	506
EXAMPLES OF PREDICTIVE ANALYTICS	508
Use of Prescriptive Analytics	509
Data Management Infrastructure for Analytics	510
Impact of Big Data and Analytics	512
Applications of Big Data and Analytics	512
BUSINESS	513
E-GOVERNMENT AND POLITICS	513
SCIENCE AND TECHNOLOGY	514
SMART HEALTH AND WELL-BEING	514
SECURITY AND PUBLIC SAFETY	514
Implications of Big Data Analytics and Decision Making	514
PERSONAL PRIVACY VS. COLLECTIVE BENEFITS	515
OWNERSHIP AND ACCESS	515
QUALITY AND REUSE OF DATA AND ALGORITHMS	515
TRANSPARENCY AND VALIDATION	516
CHANGING NATURE OF WORK	516
DEMANDS FOR WORKFORCE CAPABILITIES AND EDUCATION	516
<i>Summary</i>	516 •
<i>Key Terms</i>	517 •
<i>Review Questions</i>	517 •
<i>Problems and Exercises</i>	518 •
<i>References</i>	519 •
<i>Further Reading</i>	520 •
<i>Web Resources</i>	520

Chapter 12 Data and Database Administration 521

Learning Objectives	521
Introduction	521
The Roles of Data and Database Administrators	522
Traditional Data Administration	522
Traditional Database Administration	524
Trends in Database Administration	525
Data Warehouse Administration	527
Summary of Evolving Data Administration Roles	528

The Open Source Movement and Database Management	528
Managing Data Security	530
Threats to Data Security	531
Establishing Client/Server Security	532
SERVER SECURITY	532
NETWORK SECURITY	532
Application Security Issues in Three-Tier Client/Server Environments	533
DATA PRIVACY	534
Database Software Data Security Features	535
Views	536
Integrity Controls	536
Authorization Rules	538
User-Defined Procedures	539
Encryption	539
Authentication Schemes	540
PASSWORDS	541
STRONG AUTHENTICATION	541
Sarbanes-Oxley (SOX) and Databases	541
IT Change Management	542
Logical Access to Data	542
PERSONNEL CONTROLS	542
PHYSICAL ACCESS CONTROLS	543
IT Operations	543
Database Backup and Recovery	543
Basic Recovery Facilities	544
BACKUP FACILITIES	544
JOURNALIZING FACILITIES	544
CHECKPOINT FACILITY	545
RECOVERY MANAGER	545
Recovery and Restart Procedures	546
DISK MIRRORING	546
RESTORE/RERUN	546
MAINTAINING TRANSACTION INTEGRITY	546
BACKWARD RECOVERY	548
FORWARD RECOVERY	549
Types of Database Failure	549
ABORTED TRANSACTIONS	549
INCORRECT DATA	549
SYSTEM FAILURE	550
DATABASE DESTRUCTION	550
Disaster Recovery	550
Controlling Concurrent Access	551
The Problem of Lost Updates	551
Serializability	551
Locking Mechanisms	552
LOCKING LEVEL	552
TYPES OF LOCKS	553
DEADLOCK	554
MANAGING DEADLOCK	554

Versioning	555
Data Dictionaries and Repositories	557
Data Dictionary	557
Repositories	557
Overview of Tuning the Database for Performance	559
Installation of the DBMS	559
Memory and Storage Space Usage	559
Input/Output (I/O) Contention	560
CPU Usage	560
Application Tuning	561
Data Availability	562
Costs of Downtime	562
Measures to Ensure Availability	562
HARDWARE FAILURES	563
LOSS OR CORRUPTION OF DATA	563
HUMAN ERROR	563
MAINTENANCE DOWNTIME	563
NETWORK-RELATED PROBLEMS	563
<i>Summary</i>	<i>564</i>
<i>Key Terms</i>	<i>564</i>
<i>Review Questions</i>	<i>565</i>
<i>Problems and Exercises</i>	<i>566</i>
<i>Field Exercises</i>	<i>568</i>
<i>References</i>	<i>568</i>
<i>Further Reading</i>	<i>569</i>
<i>Web Resources</i>	<i>569</i>
<i>Glossary of Acronyms</i>	<i>570</i>
<i>Glossary of Terms</i>	<i>572</i>
<i>Index</i>	<i>580</i>

ONLINE CHAPTERS

Chapter 13 Distributed Databases 13-1

Learning Objectives 13-1

Introduction 13-1

Objectives and Trade-offs 13-4

Options for Distributing a Database 13-6

Data Replication 13-6

SNAPSHOT REPLICATION 13-7

NEAR-REAL-TIME REPLICATION 13-8

PULL REPLICATION 13-8

DATABASE INTEGRITY WITH REPLICATION 13-8

WHEN TO USE REPLICATION 13-8

Horizontal Partitioning 13-9

Vertical Partitioning 13-10

Combinations of Operations 13-11

Selecting the Right Data Distribution Strategy 13-11

Distributed DBMS 13-13

Location Transparency 13-15

Replication Transparency 13-16

Failure Transparency 13-17

Commit Protocol 13-17

Concurrency Transparency 13-18

TIME-STAMPING 13-18

Query Optimization 13-19

Evolution of Distributed DBMSs 13-21

REMOTE UNIT OF WORK 13-22

DISTRIBUTED UNIT OF WORK 13-22

DISTRIBUTED REQUEST 13-23

Summary 13-23 • Key Terms 13-24 • Review Questions 13-24 •

Problems and Exercises 13-25 • Field Exercises 13-26 •

References 13-27 • Further Reading 13-27 •

Web Resources 13-27

Chapter 14 Object-Oriented Data Modeling 14-1

Learning Objectives 14-1

Introduction 14-1

Unified Modeling Language 14-3

Object-Oriented Data Modeling 14-4

Representing Objects and Classes 14-4

Types of Operations 14-7

Representing Associations 14-7

Representing Association Classes 14-11

Representing Derived Attributes, Derived Associations,
and Derived Roles 14-12

Representing Generalization 14-13

Interpreting Inheritance and Overriding 14-18

Representing Multiple Inheritance 14-19

Representing Aggregation 14-19

Business Rules 14-22

Object Modeling Example: Pine Valley Furniture Company 14-23

[Summary](#) 14-25 • [Key Terms](#) 14-26 • [Review Questions](#) 14-26 •
[Problems and Exercises](#) 14-30 • [Field Exercises](#) 14-37 •
[References](#) 14-37 • [Further Reading](#) 14-38 •
[Web Resources](#) 14-38



Appendix A Data Modeling Tools and Notation A-1

Comparing E-R Modeling Conventions A-1

Visio Professional 2013 Notation A-1

ENTITIES A-5

RELATIONSHIPS A-5

CA ERwin Data Modeler 9.5 Notation A-5

ENTITIES A-5

RELATIONSHIPS A-5

SAP Sybase PowerDesigner 16.5 Notation A-7

ENTITIES A-8

RELATIONSHIPS A-8

Oracle Designer Notation A-8

ENTITIES A-8

RELATIONSHIPS A-8

Comparison of Tool Interfaces and E-R Diagrams A-8

Appendix B Advanced Normal Forms B-1

Boyce-Codd Normal Form B-1

Anomalies in Student Advisor B-1

Definition of Boyce-Codd Normal Form (BCNF) B-2

Converting a Relation to BCNF B-2

Fourth Normal Form B-3

Multivalued Dependencies B-5

Higher Normal Forms B-5

[Key Terms](#) B-6 • [References](#) B-6 • [Web Resource](#) B-6

Appendix C Data Structures C-1

Pointers C-1

Data Structure Building Blocks C-2

Linear Data Structures C-4

Stacks C-5

Queues C-5

Sorted Lists C-6

Multilists C-8

Hazards of Chain Structures C-8

Trees C-9

Balanced Trees C-9

[Reference](#) C-12

This page intentionally left blank

PREFACE

This text is designed to be used with an introductory course in database management. Such a course is usually required as part of an information systems curriculum in business schools, computer technology programs, and applied computer science departments. The Association for Information Systems (AIS), the Association for Computing Machinery (ACM), and the International Federation of Information Processing Societies (IFIPS) curriculum guidelines (e.g., IS 2010) all outline this type of database management course. Previous editions of this text have been used successfully for more than 33 years at both the undergraduate and graduate levels, as well as in management and professional development programs.

WHAT'S NEW IN THIS EDITION?

This 12th edition of *Modern Database Management* updates and expands materials in areas undergoing rapid change as a result of improved managerial practices, database design tools and methodologies, and database technology. Later, we detail changes to each chapter. The themes of this 12th edition reflect the major trends in the information systems field and the skills required of modern information systems graduates:

- Given the explosion in interest in the topics of big data and analytics, we have added an entire new chapter (Chapter 11) dedicated to this area. The chapter provides in-depth coverage of big data technologies such as NoSQL, Hadoop, MapReduce, Pig, and Hive and provides an introduction to the different types of analytics (descriptive, predictive, and prescriptive) and their use in business.
- We have also introduced this topic in relevant places throughout the textbook, e.g., in the revised introduction section in Chapter 1 as well as in a new section titled “The Future of Data Warehousing: Integration with Big Data and Analytics” in the data warehousing chapter (Chapter 9).
- Topics such as in-memory databases, in-database analytics, data warehousing in the cloud, and massively parallel processing are covered in sections of Chapter 9 and Chapter 11.
- The Mountain View Community Hospital (MVCH) case study (a staple of many past editions) has been replaced with a simpler mini-case titled “Forondo Artist Management Excellence Inc.” (FAME). The case focuses on the development of a system to support the needs of a small artist management company. The case is presented in the form of stakeholder e-mails describing the current challenges faced by the organization as well as the features they would like to see in a new system. Each chapter presents a set of project exercises that serve as guidelines for deliverables for students.
- We have updated the section on routines in Chapter 7 to provide clarity on the nature of routines and how to use them.
- New material added to Chapter 2 on why data modeling is important provides several compelling reasons for why data modeling is still crucial.

In addition to the new topics covered, specific improvements to the textbook have been made in the following areas:

- Every chapter went through significant edits to streamline coverage to ensure relevance with current technologies and eliminate redundancies.
- End-of-chapter material (review questions, problems and exercises, and/or field exercises) in every chapter has been revised with new questions and exercises.
- The figures in several chapters were updated to reflect the changing landscape of technologies that are being used in modern organizations.
- The Web Resources section in each chapter was updated to ensure that the student has information on the latest database trends and expanded background details on important topics covered in the text.



- We have continued to focus on reducing the length of the printed book, an effort that began with the eighth edition. The reduced length is more consistent with what our reviewers say can be covered in a database course today, given the need for depth of coverage in the most important topics. The reduced length should encourage more students to purchase and read the text, without any loss of coverage and learning. The book continues to be available through CourseSmart, an innovative e-book delivery system, and as an electronic book in the Kindle format.

Also, we continue to provide on the student Companion Web site several custom-developed short videos that address key concepts and skills from different sections of the book. These videos, produced by the textbook authors, help students learn difficult material by using both the printed text and a mini lecture or tutorial. Videos have been developed to support Chapters 1 (introduction to database), 2 and 3 (conceptual data modeling), 4 (normalization), and 6 and 7 (SQL). More will be produced with future editions. Look for special icons on the opening page of these chapters to call attention to these videos, and go to www.pearsonhighered.com/hoffer to find these videos.

FOR THOSE NEW TO *MODERN DATABASE MANAGEMENT*

Modern Database Management has been a leading text since its first edition in 1983. In spite of this market leadership position, some instructors have used other good database management texts. Why might you want to switch at this time? There are several good reasons:

- One of our goals, in every edition, has been to lead other books in coverage of the latest principles, concepts, and technologies. See what we have added for the 12th edition in “What’s New in This Edition?” In the past, we have led in coverage of object-oriented data modeling and UML, Internet databases, data warehousing, and the use of CASE tools in support of data modeling. For the 12th edition, we continue this tradition by providing significant coverage on the important topic of big data and analytics, focusing on what every database student needs to understand about these topics.
- While remaining current, this text focuses on what leading practitioners say is most important for database developers. We work with many practitioners, including the professionals of the Data Management Association (DAMA) and The Data Warehousing Institute (TDWI), leading consultants, technology leaders, and authors of articles in the most widely read professional publications. We draw on these experts to ensure that what the book includes is important and covers not only important entry-level knowledge and skills, but also those fundamentals and mind-sets that lead to long-term career success.
- In the 12th edition of this highly successful book, material is presented in a way that has been viewed as very accessible to students. Our methods have been refined through continuous market feedback for more than 30 years, as well as through our own teaching. Overall, the pedagogy of the book is sound. We use many illustrations that help make important concepts and techniques clear. We use the most modern notations. The organization of the book is flexible, so you can use chapters in whatever sequence makes sense for your students. We supplement the book with data sets to facilitate hands-on, practical learning, and with new media resources to make some of the more challenging topics more engaging.
- Our text can accommodate structural flexibility. For example, you may have particular interest in introducing SQL early in your course. Our text makes this possible. First, we cover SQL in depth, devoting two full chapters to this core technology of the database field. Second, we include many SQL examples in early chapters. Third, many instructors have successfully used the two SQL chapters early in their course. Although logically appearing in the life cycle of systems development as Chapters 6 and 7, part of the implementation section of the text, many instructors have used these chapters immediately after Chapter 1 or in parallel with other early chapters. Finally, we use SQL throughout the book, for example, to illustrate Web application connections to relational databases in Chapter 8 and online analytical processing in Chapter 11.

- We have the latest in supplements and Web site support for the text. See the supplement package for details on all the resources available to you and your students.
- This text is written to be part of a modern information systems curriculum with a strong business systems development focus. Topics are included and addressed so as to reinforce principles from other typical courses, such as systems analysis and design, networking, Web site design and development, MIS principles, and computer programming. Emphasis is on the development of the database component of modern information systems and on the management of the data resource. Thus, the text is practical, supports projects and other hands-on class activities, and encourages linking database concepts to concepts being learned throughout the curriculum the student is taking.

SUMMARY OF ENHANCEMENTS TO EACH CHAPTER

The following sections present a chapter-by-chapter description of the major changes in this edition. Each chapter description presents a statement of the purpose of that chapter, followed by a description of the changes and revisions that have been made for the 12th edition. Each paragraph concludes with a description of the strengths that have been retained from prior editions.

PART I: THE CONTEXT OF DATABASE MANAGEMENT

Chapter 1: The Database Environment and Development Process

This chapter discusses the role of databases in organizations and previews the major topics in the remainder of the text. The primary change in this chapter has been in how we use current examples around the explosion in the amount of data being generated and the benefits that can be gained by harnessing the power data (through analytics) to help set the stage for the entire book. A few new exercises have also been added, and the new Forondo Artist Management Excellence (FAME) case is introduced. After presenting a brief introduction to the basic terminology associated with storing and retrieving data, the chapter presents a well-organized comparison of traditional file processing systems and modern database technology. The chapter then introduces the core components of a database environment. It then goes on to explain the process of database development in the context of structured life cycle, prototyping, and agile methodologies. The presentation remains consistent with the companion textbook, *Modern Systems Analysis and Design* by Hoffer, George, and Valacich. The chapter also discusses important issues in database development, including management of the diverse group of people involved in database development and frameworks for understanding database architectures and technologies (e.g., the three-schema architecture). Reviewers frequently note the compatibility of this chapter with what students learn in systems analysis and design classes. A brief history of the evolution of database technology, from pre-database files to modern object-relational technologies, is presented. The chapter also provides an overview of the range of database applications that are currently in use within organizations—personal, two-tier, multitier, and enterprise applications. The explanation of enterprise databases includes databases that are part of enterprise resource planning systems and data warehouses. The chapter concludes with a description of the process of developing a database in a fictitious company, Pine Valley Furniture. This description closely mirrors the steps in database development described earlier in the chapter.

PART II: DATABASE ANALYSIS

Chapter 2: Modeling Data in the Organization

This chapter presents a thorough introduction to conceptual data modeling with the entity-relationship (E-R) model. The chapter title emphasizes the reason for the entity-relationship model: to unambiguously document the rules of the business that influence database design. New material on why data modeling is important helps set the stage for the rest of the discussion that follows. Specific subsections explain in detail how to name and define elements of a data model, which are essential in

developing an unambiguous E-R diagram. The chapter continues to proceed from simple to more complex examples, and it concludes with a comprehensive E-R diagram for the Pine Valley Furniture Company. In the 12th edition, we have provided three new problems and exercises, and the second part of the new FAME case is introduced. Appendix A provides information on different data modeling tools and notations.

Chapter 3: The Enhanced E-R Model

This chapter presents a discussion of several advanced E-R data model constructs, primarily supertype/subtype relationships. As in Chapter 2, problems and exercises have been revised. The third part of the new FAME case is presented in this chapter. The chapter continues to present thorough coverage of supertype/subtype relationships and includes a comprehensive example of an extended E-R data model for the Pine Valley Furniture Company.

PART III: DATABASE DESIGN

Chapter 4: Logical Database Design and the Relational Model

This chapter describes the process of converting a conceptual data model to the relational data model, as well as how to merge new relations into an existing normalized database. It provides a conceptually sound and practically relevant introduction to normalization, emphasizing the importance of the use of functional dependencies and determinants as the basis for normalization. Concepts of normalization and normal forms are extended in Appendix B. The chapter features a discussion of the characteristics of foreign keys and introduces the important concept of a nonintelligent enterprise key. Enterprise keys (also called surrogate keys for data warehouses) are emphasized as some concepts of object orientation have migrated into the relational technology world. Eight new review questions and problems and exercises are included, and the revision has further clarified the coverage of some of the key concepts and the visual quality of the presentation. The chapter continues to emphasize the basic concepts of the relational data model and the role of the database designer in the logical design process. The new FAME case continues in this chapter.

Chapter 5: Physical Database Design and Performance

This chapter describes the steps that are essential in achieving an efficient database design, with a strong focus on those aspects of database design and implementation that are typically within the control of a database professional in a modern database environment. Five new review questions and problems and exercises are included. In addition, the language of the chapter was streamlined to improve readability. References to Oracle (including the visual coverage of database terminology) were updated to cover the latest version (at the time of this writing), 12c. New coverage of heap file organization was added to the chapter. The chapter contains an emphasis on ways to improve database performance, with references to specific techniques available in Oracle and other DBMSs to improve database processing performance. The discussion of indexes includes descriptions of the types of indexes (primary and secondary indexes, join index, hash index table) that are widely available in database technologies as techniques to improve query processing speed. Appendix C provides excellent background on fundamental data structures for programs of study that need coverage of this topic. The chapter continues to emphasize the physical design process and the goals of that process. The new FAME case continues with questions related to the material covered in this chapter.

PART IV: IMPLEMENTATION

Chapter 6: Introduction to SQL

This chapter presents a thorough introduction to the SQL used by most DBMSs (SQL:1999) and introduces the changes that are included in the latest standard (SQL:2011). This edition adds coverage of the new features of SQL:2011. The coverage of SQL is extensive

and divided into this and the next chapter. This chapter includes examples of SQL code, using mostly SQL:1999 and SQL:2011 syntax, as well as some Oracle 12c and Microsoft SQL Server syntax. Some unique features of MySQL are mentioned. Both dynamic and materialized views are also covered. This revision links Chapter 6 explicitly with the material covered in the new Chapter 11 on big data and analytics. Chapter 6 explains the SQL commands needed to create and maintain a database and to program single-table queries. The revised version of the chapter provides the reader with improved guidance regarding alternate sequences for learning the material. Coverage of dual-table, IS NULL/IS NOT NULL, more built-in functions, derived tables, and rules for aggregate functions and the GROUP BY clause is included or improved. Three review questions and eight problems and exercises have been added to the chapter. The chapter continues to use the Pine Valley Furniture Company case to illustrate a wide variety of practical queries and query results. Questions related to the new FAME case also are available in the context of this chapter.

Chapter 7: Advanced SQL

This chapter continues the description of SQL, with a careful explanation of multiple-table queries, transaction integrity, data dictionaries, triggers and stored procedures (the differences between them are now more clearly explained), and embedded SQL in other programming language programs. All forms of the OUTER JOIN command are covered. Standard SQL (with an updated focus on SQL:2011) is also used. The revised version of the chapter includes a new section on the temporal features introduced in SQL:2011. This chapter illustrates how to store the results of a query in a derived table, the CAST command to convert data between different data types, and the CASE command for doing conditional processing in SQL. Emphasis continues on the set-processing style of SQL compared with the record processing of programming languages with which the student may be familiar. The section on routines has been revised to provide clarified, expanded, and more current coverage of this topic. New and updated problems and exercises have been added to the chapter. The chapter continues to contain a clear explanation of subqueries and correlated subqueries, two of the most complex and powerful constructs in SQL. This chapter also includes relevant FAME case questions.

Chapter 8: Database Application Development

This chapter provides a modern discussion of the concepts of client/server architecture and applications, middleware, and database access in contemporary database environments. The section has been revised to ensure that the applicability of the concepts presented in the chapter is clear in the era of modern devices such as smartphones, tablets, etc. Review questions and problems and exercises have been updated. The chapter focuses on technologies that are commonly used to create two- and three-tier applications. Many figures are included to show the options in multitiered networks, including application and database servers, database processing distribution alternatives among network tiers, and browser (thin) clients. The chapter also presents sample application programs that demonstrate how to access databases from popular programming languages such as Java, VB.NET, ASP.NET, JSP, and PHP. This chapter lays the technology groundwork for the Internet topics presented in the remainder of the text and highlights some of the key considerations in creating three-tier Internet-based applications. The chapter also provides coverage of the role of Extensible Markup Language (XML) and related technologies in data storage and retrieval. Topics covered include basics of XML schemas, XQuery, and XSLT. The chapter concludes with an overview of Web services; associated standards and technologies; and their role in seamless, secure movement of data in Web-based applications. A brief introduction to service-oriented architecture (SOA) is also presented. Security topics, including Web security, are covered in Chapter 12. This chapter includes the final questions related to the new FAME case.

Chapter 9: Data Warehousing

This chapter describes the basic concepts of data warehousing, the reasons data warehousing is regarded as critical to competitive advantage in many organizations, and the database design activities and structures unique to data warehousing. A new section on

the future of data warehousing provides a preview of the topics that will be covered in the new chapter (Chapter 11) on big data and analytics and serves as the link between these two chapters. Some of the material that previously belonged to this chapter is now covered in an expanded fashion in Chapter 11. Topics covered in this chapter include alternative data warehouse architectures and the dimensional data model (or star schema) for data warehouses. Coverage of architectures has been streamlined consistent with trends in data warehousing, and a deep explanation of how to handle slowly changing dimensional data is provided. Operational data store and independent, dependent, and logical data marts are defined.

PART V: ADVANCED DATABASE TOPICS

Chapter 10: Data Quality and Integration

In this chapter, the principles of data governance, which are at the core of enterprise data management (EDM) activities, are introduced. This is followed by coverage of data quality. This chapter describes the need for an active program to manage data quality in organizations and outlines the steps that are considered today to be best practices for data quality management. Quality data are defined, and reasons for poor-quality data are identified. Methods for data quality improvement, such as data auditing, improving data capturing (a key part of database design), data stewardship and governance, TQM principles, modern data management technologies, and high-quality data models are all discussed. The topic of master data management, one approach to integrating key business data, is introduced and explained. Different approaches to data integration are overviewed, and the reasons for each are outlined. The extract, transform, load (ETL) process for data warehousing is discussed in detail.

Chapter 11: Big Data and Analytics

Chapter 11 on big data and analytics is new in this edition, and it extends the coverage of the text in three important ways: First, this chapter provides a systematic introduction to the technologies that are currently discussed under the label *big data* and the impact of these technologies on the overall enterprise data management architecture. Specifically, the chapter focuses on the Hadoop infrastructure and four categories of so-called NoSQL (Not only SQL) database management systems. Second, the chapter offers integrated coverage of analytics, including descriptive, predictive, and prescriptive analytics. The discussion on analytics is linked not only to the coverage of big data but also the material on data warehousing in Chapter 9 and the general discussion on data management in Chapter 1. The chapter also briefly covers approaches and technologies used by analytics professionals, such as OLAP, data visualization, business performance management and dashboards, data mining, and text mining. Third, the chapter integrates the coverage of big data and analytics technologies to the individual, organizational, and societal implications of these capabilities.

Chapter 12: Data and Database Administration

This chapter presents a thorough discussion of the importance and roles of data and database administration and describes a number of the key issues that arise when these functions are performed. This chapter emphasizes the changing roles and approaches of data and database administration, with emphasis on data quality and high performance. We also briefly touch upon the impact of cloud computing on the data/database administration. The chapter contains a thorough discussion of database backup procedures, as well as extensively expanded and consolidated coverage of data security threats and responses and data availability. The data security topics include database security policies, procedures, and technologies (including encryption and smart cards). The role of databases in Sarbanes-Oxley compliance is also examined. We also discuss open source DBMS, the benefits and hazards of this technology, and how to choose an open source DBMS. In addition, the topic of heartbeat queries is included in the coverage of database performance improvements. The chapter continues to emphasize the critical importance of data and database management in managing data as a corporate asset.