

GLOBAL
EDITION



Modern Database Management

THIRTEENTH EDITION

Jeffrey A. Hoffer
V. Ramesh
Heikki Topi



THIRTEENTH EDITION
GLOBAL EDITION

MODERN DATABASE MANAGEMENT

Jeffrey A. Hoffer
University of Dayton

V. Ramesh
Indiana University

Heikki Topi
Bentley University



Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Vice President, IT & Careers: Andrew Gilfillan
Senior Portfolio Manager: Samantha Lewis
Managing Producer: Laura Burgess
Associate Content Producer: Stephany Harrington
Content Producer, Global Edition: Sonam Arora
Assistant Acquisitions Editor, Global Edition: Rosemary Iles
Senior Project Editor, Global Edition: Daniel Luiz
Manager, Media Production, Global Edition: Gargi Banerjee
Manufacturing Controller, Production, Global Edition: Kay Holman
Portfolio Management Assistant: Madeline Houpt
Director of Product Marketing: Brad Parkins
Product Marketing Manager: Heather Taylor

Product Marketing Assistant: Jesika Bethea
Field Marketing Manager: Molly Schmidt
Field Marketing Assistant: Kelli Fisher
Cover Image: mistery/Shutterstock
Vice President, Product Model Management: Jason Fournier
Senior Product Model Manager: Eric Hakanson
Lead, Production and Digital Studio: Heather Darby
Digital Studio Course Producer: Jaimie Noy
Program Monitor: Danica Monzor, SPi Global
Full-Service Project Management: Neha Bhargava, Cenveo® Publisher Services
Composition: Cenveo Publisher Services

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 noninfringement. 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

KAO Two
KAO Park
Harlow
CM17 9NA
United Kingdom

and Associated Companies throughout the world

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

© Pearson Education Limited 2020

The rights of Jeffrey A. Hoffer, V. Ramesh, 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, 13th edition, ISBN 978-0-13-477365-0, by Jeffrey A. Hoffer, V. Ramesh, and Heikki Topi, published by Pearson Education © 2019.

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 EC1N 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: 1-292-26335-0

ISBN 13: 978-1-292-26335-9

eBook ISBN: 978-1-292-26341-0

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 in Palatino LT Pro by Cenveo® Publisher Services

To Patty, for her sacrifices, encouragement, and support for more than 35 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 35

Chapter 1 The Database Environment and Development Process 37

Part II Database Analysis and Logical Design 87

Chapter 2 Modeling Data in the Organization 89

Chapter 3 The Enhanced E-R Model 149

Chapter 4 Logical Database Design and the Relational Model 187

Part III Database Implementation and Use 239

Chapter 5 Introduction to SQL 241

Chapter 6 Advanced SQL 285

Chapter 7 Databases in Applications 331

Chapter 8 Physical Database Design and Database Infrastructure 367

Part IV Advanced Database Topics 419

Chapter 9 Data Warehousing and Data Integration 421

Chapter 10 Big Data Technologies 478

Chapter 11 Analytics and Its Implications 508

Chapter 12 Data and Database Administration with Focus on Data Quality 537

Glossary of Acronyms 563

Glossary of Terms 565

Index 573

Available Online at www.pearsonglobaleditions.com

Chapter 13 Distributed Databases 13-1

Chapter 14 Object-Oriented Data Modeling 14-1

Appendices

Appendix A Data Modeling Tools and Notation A-1

Appendix B Advanced Normal Forms B-1

Appendix C Data Structures C-1

This page intentionally left blank

CONTENTS

Preface 23

Part I The Context of Database Management 35

An Overview of Part I 35

Chapter 1 The Database Environment and Development Process 37



Learning Objectives 37

Data Matter! 38

Introduction 39

Basic Concepts and Definitions 40

Data 40

Data versus Information 41

Metadata 42

Traditional File Processing Systems 43

File Processing Systems at Pine Valley Furniture Company 43

Disadvantages of File Processing Systems 44

PROGRAM-DATA DEPENDENCE 44

DUPPLICATION OF DATA 44

LIMITED DATA SHARING 44

LENGTHY DEVELOPMENT TIMES 44

EXCESSIVE PROGRAM MAINTENANCE 45

The Database Approach 45

Data Models 45

ENTITIES 45

RELATIONSHIPS 45

Relational Databases 46

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 48

ENFORCEMENT OF STANDARDS 49

IMPROVED DATA QUALITY 49

IMPROVED DATA ACCESSIBILITY AND RESPONSIVENESS 49

REDUCED PROGRAM MAINTENANCE 50

IMPROVED DECISION SUPPORT 50

CAUTIONS ABOUT DATABASE BENEFITS 50

COSTS AND RISKS OF THE DATABASE APPROACH 50

NEW, SPECIALIZED PERSONNEL 50

INSTALLATION AND MANAGEMENT COST AND COMPLEXITY 51

CONVERSION COSTS 51

NEED FOR EXPLICIT BACKUP AND RECOVERY 51

ORGANIZATIONAL CONFLICT 51

Integrated Data Management Framework 51

Components of the Database Environment 52

The Database Development Process	54
Systems Development Life Cycle	55
PLANNING—ENTERPRISE MODELING	55
PLANNING—CONCEPTUAL DATA MODELING	55
ANALYSIS—CONCEPTUAL DATA MODELING	56
DESIGN—LOGICAL DATABASE DESIGN	57
DESIGN—PHYSICAL DATABASE DESIGN AND DEFINITION	57
IMPLEMENTATION—DATABASE IMPLEMENTATION	57
MAINTENANCE—DATABASE MAINTENANCE	58
Alternative Information Systems Development Approaches	58
Three-Schema Architecture for Database Development	59
Managing the People Involved in Database Development	61
Evolution of Database Systems	61
1960s	63
1970s	63
1980s	63
1990s	64
2000 and Beyond	64
The Range of Database Applications	64
Personal Databases	65
Departmental Multi-Tiered Client/Server Databases	65
Enterprise Applications	66
ENTERPRISE SYSTEMS	66
DATA WAREHOUSES	67
DATA LAKE	68
Developing a Database Application for Pine Valley Furniture Company	69
Database Evolution at Pine Valley Furniture Company	70
Project Planning	70
Analyzing Database Requirements	71
Designing the Database	74
Using the Database	76
Administering the Database	77
Future of Databases at Pine Valley	77
<i>Summary</i>	78
<i>Key Terms</i>	79
<i>Review Questions</i>	79
<i>Problems and Exercises</i>	80
<i>Field Exercises</i>	82
<i>References</i>	83
<i>Further Reading</i>	83
<i>Web Resources</i>	84
▶ <i>CASE: Forondo Artist Management Excellence Inc.</i>	85



Part II Database Analysis and Logical Design 87

An Overview of Part II 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 115
 - BINARY RELATIONSHIP 116
 - TERNARY RELATIONSHIP 116
 - Attributes or Entity? 117
 - Cardinality Constraints 119
 - MINIMUM CARDINALITY 119
 - MAXIMUM CARDINALITY 120
 - Some Examples of Relationships and Their Cardinalities 120
 - A TERNARY RELATIONSHIP 121
 - Modeling Time-Dependent Data 122
 - Modeling Multiple Relationships Between Entity Types 124
 - Naming and Defining Relationships 126
 - E-R Modeling Example: Pine Valley Furniture Company 127
 - Database Processing At Pine Valley Furniture 130
 - Showing Product Information 130
 - Showing Product Line Information 130
 - Showing Customer Order Status 131
 - Showing Product Sales 132
 - Summary 133 • Key Terms 134 • Review Questions 134 • Problems and Exercises 135 • Field Exercises 145 • References 146 • Further Reading 146 • Web Resources 146*
 - ▶ **CASE: Forondo Artist Management Excellence Inc. 147**





Chapter 3 The Enhanced E-R Model 149

Learning Objectives	149
Introduction	149
Representing Supertypes and Subtypes	150
Basic Concepts and Notation	151
AN EXAMPLE OF A SUPERTYPE/SUBTYPE RELATIONSHIP	152
ATTRIBUTE INHERITANCE	153
WHEN TO USE SUPERTYPE/SUBTYPE RELATIONSHIPS	153
Representing Specialization and Generalization	154
GENERALIZATION	154
SPECIALIZATION	155
COMBINING SPECIALIZATION AND GENERALIZATION	156
Specifying Constraints in Supertype/Subtype Relationships	157
Specifying Completeness Constraints	157
TOTAL SPECIALIZATION RULE	157
PARTIAL SPECIALIZATION RULE	157
Specifying Disjointness Constraints	158
DISJOINT RULE	158
OVERLAP RULE	159
Defining Subtype Discriminators	159
DISJOINT SUBTYPES	159
OVERLAPPING SUBTYPES	160
Defining Supertype/Subtype Hierarchies	161
AN EXAMPLE OF A SUPERTYPE/SUBTYPE HIERARCHY	162
SUMMARY OF SUPERTYPE/SUBTYPE HIERARCHIES	162
EER Modeling Example: Pine Valley Furniture Company	162
Entity Clustering	166
Packaged Data Models	169
A Revised Data Modeling Process with Packaged Data Models	171
Packaged Data Model Examples	173
<i>Summary</i>	178
<i>Key Terms</i>	179
<i>Review Questions</i>	179
<i>Problems and Exercises</i>	180
<i>Field Exercises</i>	182
<i>References</i>	183
<i>Further Reading</i>	183
<i>Web Resources</i>	183
▶ CASE: Forondo Artist Management Excellence Inc.	185



Chapter 4 Logical Database Design and the Relational Model 187

Learning Objectives	187
Introduction	187
The Relational Data Model	188
Basic Definitions	188
RELATIONAL DATA STRUCTURE	189
RELATIONAL KEYS	189
PROPERTIES OF RELATIONS	190
REMOVING MULTIVALUED ATTRIBUTES FROM TABLES	190
Sample Database	191
Integrity Constraints	192
Domain Constraints	192
Entity Integrity	192
Referential Integrity	194

Creating Relational Tables	195
Well-Structured Relations	196
Transforming EER Diagrams into Relations	197
Step 1: Map Regular Entities	198
COMPOSITE ATTRIBUTES	198
MULTIVALUED ATTRIBUTES	199
Step 2: Map Weak Entities	199
WHEN TO CREATE A SURROGATE KEY	200
Step 3: Map Binary Relationships	201
MAP BINARY ONE-TO-MANY RELATIONSHIPS	201
MAP BINARY MANY-TO-MANY RELATIONSHIPS	202
MAP BINARY ONE-TO-ONE RELATIONSHIPS	202
Step 4: Map Associative Entities	203
IDENTIFIER NOT ASSIGNED	203
IDENTIFIER ASSIGNED	204
Step 5: Map Unary Relationships	205
UNARY ONE-TO-MANY RELATIONSHIPS	205
UNARY MANY-TO-MANY RELATIONSHIPS	206
Step 6: Map Ternary (and n -ary) Relationships	207
Step 7: Map Supertype/Subtype Relationships	208
Summary of EER-to-Relational Transformations	210
Introduction to Normalization	210
Steps in Normalization	211
Functional Dependencies and Keys	211
DETERMINANTS	213
CANDIDATE KEYS	213
Normalization Example: Pine Valley Furniture Company	214
Step 0: Represent the View in Tabular Form	214
Step 1: Convert to First Normal Form	215
REMOVE REPEATING GROUPS	215
SELECT THE PRIMARY KEY	216
ANOMALIES IN 1NF	216
Step 2: Convert to Second Normal Form	217
Step 3: Convert to Third Normal Form	218
REMOVING TRANSITIVE DEPENDENCIES	218
Determinants and Normalization	219
Step 4: Further Normalization	219
Merging Relations	220
An Example	220
View Integration Problems	220
SYNONYMS	221
HOMONYMS	221
TRANSITIVE DEPENDENCIES	221
SUPERTYPE/SUBTYPE RELATIONSHIPS	222
A Final Step for Defining Relational Keys	222
<i>Summary</i>	225
<i>Key Terms</i>	225
<i>Review Questions</i>	225
<i>Problems and Exercises</i>	226
<i>Field Exercises</i>	235
<i>References</i>	235
<i>Further Reading</i>	236
<i>Web Resources</i>	236
▶ CASE: Forondo Artist Management Excellence Inc.	237



Part III Database Implementation and Use 239

An Overview of Part III 239



Chapter 5 Introduction to SQL 241

Learning Objectives 241

Introduction 241

Origins of the SQL Standard 243

The SQL Environment 245

SQL Data Types 247

Defining A Database in SQL 250

Generating SQL Database Definitions 250

Creating Tables 251

Creating Data Integrity Controls 254

Changing Table Definitions 255

Removing Tables 255

Inserting, Updating, and Deleting Data 256

Batch Input 257

Deleting Database Contents 257

Updating Database Contents 258

Internal Schema Definition in RDBMSs 259

Creating Indexes 259

Processing Single Tables 260

Clauses of the SELECT Statement 260

Using Expressions 262

Using Functions 263

Using Wildcards 266

Using Comparison Operators 266

Using Null Values 267

Using Boolean Operators 267

Using Ranges for Qualification 270

Using Distinct Values 270

Using IN and NOT IN with Lists 272

Sorting Results: The ORDER BY Clause 273

Categorizing Results: The GROUP BY Clause 274

Qualifying Results by Categories: The HAVING Clause 275

Summary 277 • Key Terms 277 • Review Questions 277 •

Problems and Exercises 278 • Field Exercises 282 •

References 282 • Further Reading 283 •

Web Resources 283

▶ **CASE: Forondo Artist Management Excellence Inc. 284**



Chapter 6 Advanced SQL 285

Learning Objectives 285

Introduction 285

Processing Multiple Tables 286

Equi-Join 287

Natural Join 288

Outer Join 289

Sample Join Involving Four Tables 291

Self-Join	292
Subqueries	294
Correlated Subqueries	299
Using Derived Tables	301
Combinings Queries	301
Conditional Expressions	303
More Complicated SQL Queries	304
Tips for Developing Queries	306
Guidelines for Better Query Design	308
Using and Defining Views	309
Materialized Views	313
Triggers and Routines	313
Triggers	314
Routines and Other Programming Extensions	316
Example Routine in Oracle's PL/SQL	318
Data Dictionary Facilities	319
Recent Enhancements and Extensions to SQL	321
Analytical and OLAP Functions	321
New Temporal Features in SQL	322
Other Enhancements	322
<i>Summary</i>	<i>323</i>
<i>Key Terms</i>	<i>324</i>
<i>Review Questions</i>	<i>324</i>
<i>Problems and Exercises</i>	<i>325</i>
<i>Field Exercises</i>	<i>328</i>
<i>References</i>	<i>328</i>
<i>Further Reading</i>	<i>329</i>
<i>Web Resources</i>	<i>329</i>
▶ CASE: Forondo Artist Management Excellence Inc.	330



Chapter 7 Databases in Applications 331

Learning Objectives	331
Location, Location, Location!	331
Introduction	332
Client/Server Architectures	332
Databases in Three-Tier Applications	336
A Java Web Application	337
A Python Web Application	341
Key Considerations in Three-Tier Applications	347
Stored Procedures	347
Transactions	347
Database Connections	349
Key Benefits of Three-Tier Applications	349
Transaction Integrity	350
Controlling Concurrent Access	352
The Problem of Lost Updates	352
Serializability	353
Locking Mechanisms	353
LOCKING LEVEL	353
TYPES OF LOCKS	354
DEADLOCK	355
MANAGING DEADLOCK	355
Versioning	356



Managing Data Security in an Application Context	358
Threats to Data Security	358
Establishing Client/Server Security	359
SERVER SECURITY	360
NETWORK SECURITY	360
Application Security Issues in Three-Tier Client/Server Environments	360
DATA PRIVACY	361
<i>Summary</i>	363 • <i>Key Terms</i> 363 • <i>Review Questions</i> 363 • <i>Problems and Exercises</i> 364 • <i>Field Exercises</i> 364 • <i>References</i> 365 • <i>Further Reading</i> 365 • <i>Web Resources</i> 365
▶ CASE: Forondo Artist Management Excellence Inc.	366

Chapter 8 Physical Database Design and Database Infrastructure 367

Learning Objectives	367
Introduction	368
The Physical Database Design Process	369
Who Is Responsible for Physical Database Design?	369
Physical Database Design as a Basis for Regulatory Compliance	370
SOX and Databases	371
IT CHANGE MANAGEMENT	371
LOGICAL ACCESS TO DATA	371
IT OPERATIONS	372
Data Volume and Usage Analysis	372
Designing Fields	374
Choosing Data Types	374
CODING TECHNIQUES	375
CONTROLLING DATA INTEGRITY	376
HANDLING MISSING DATA	377
Denormalizing and Partitioning Data	377
Denormalization	377
OPPORTUNITIES FOR AND TYPES OF DENORMALIZATION	378
DENORMALIZE WITH CAUTION	379
Partitioning	381
Designing Physical Database Files	382
File Organizations	384
HEAP FILE ORGANIZATION	384
SEQUENTIAL FILE ORGANIZATIONS	384
INDEXED FILE ORGANIZATIONS	386
HASHED FILE ORGANIZATIONS	387
Clustering Files	387
Designing Controls for Files	388
Using and Selecting Indexes	388
Creating a Unique Key Index	388
Creating a Secondary (Nonunique) Key Index	389
When to Use Indexes	389
Designing a Database for Optimal Query Performance	390
Parallel Query Processing	391
Overriding Automatic Query Optimization	392
Data Dictionaries and Repositories	392

Data Dictionary	393
Repositories	393
Database Software Data Security Features	395
Views	395
Integrity Controls	396
Authorization Rules	397
User-Defined Procedures	399
Encryption	399
Authentication Schemes	399
PASSWORDS	400
STRONG AUTHENTICATION	400
Database Backup and Recovery	401
Basic Recovery Facilities	401
BACKUP FACILITIES	401
JOURNALIZING FACILITIES	402
CHECKPOINT FACILITY	402
RECOVERY MANAGER	403
Recovery and Restart Procedures	403
DISK MIRRORING	403
RESTORE/RERUN	404
BACKWARD RECOVERY	404
FORWARD RECOVERY	405
Types of Database Failure	405
ABORTED TRANSACTIONS	406
INCORRECT DATA	406
SYSTEM FAILURE	406
DATABASE DESTRUCTION	406
Disaster Recovery	407
Cloud-Based Database Infrastructure	407
Cloud-Based Models for Providing Data Management Services	407
Benefits and Downsides of Using Cloud-Based Data Management Services	408
<i>Summary</i>	409
<i>Key Terms</i>	410
<i>Review Questions</i>	411
<i>Problems and Exercises</i>	412
<i>Field Exercises</i>	416
<i>References</i>	417
<i>Further Reading</i>	417
<i>Web Resources</i>	417
▶ CASE: Forondo Artist Management Excellence Inc.	418



Part IV Advanced Database Topics 419

An Overview of Part IV 419

Chapter 9 Data Warehousing and Data Integration 421

Learning Objectives 421

Introduction 421

Basic Concepts of Data Warehousing 424

 A Brief History of Data Warehousing 424

 The Need for Data Warehousing 424

 NEED FOR A COMPANY-WIDE VIEW 424

 NEED TO SEPARATE OPERATIONAL AND INFORMATIONAL SYSTEMS 427

Data Warehouse Architectures 427

 Independent Data Mart Data Warehousing Environment 428

Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach	429
Logical Data Mart and Real-Time Data Warehouse Architecture	431
Three-Layer Data Architecture	434
ROLE OF THE ENTERPRISE DATA MODEL	434
ROLE OF METADATA	434
Some Characteristics of Data Warehouse Data	435
Status versus Event Data	435
Transient versus Periodic Data	436
An Example of Transient and Periodic Data	436
TRANSIENT DATA	438
PERIODIC DATA	438
OTHER DATA WAREHOUSE CHANGES	438
The Derived Data Layer	439
Characteristics of Derived Data	439
The Star Schema	440
FACT TABLES AND DIMENSION TABLES	440
EXAMPLE STAR SCHEMA	441
SURROGATE KEY	442
GRAIN OF THE FACT TABLE	443
DURATION OF THE DATABASE	444
SIZE OF THE FACT TABLE	444
MODELING DATE AND TIME	445
Variations of the Star Schema	446
MULTIPLE FACT TABLES	446
FACTLESS FACT TABLES	447
Normalizing Dimension Tables	448
MULTIVALUED DIMENSIONS	448
HIERARCHIES	449
Slowly Changing Dimensions	451
Determining Dimensions and Facts	454
Data Integration: An Overview	456
General Approaches to Data Integration	456
DATA FEDERATION	457
DATA PROPAGATION	457
Data Integration for Data Warehousing: The Reconciled Data Layer	458
Characteristics of Data after ETL	458
The ETL Process	459
MAPPING AND METADATA MANAGEMENT	459
EXTRACT	460
CLEANSE	461
LOAD AND INDEX	463
Data Transformation	464
Data Transformation Functions	465
RECORD-LEVEL FUNCTIONS	465
FIELD-LEVEL FUNCTIONS	466
Data Warehouse Administration	468

The Future of Data Warehousing: Integration with Other Forms of Data Management and Analytics	468
Speed of Processing	469
Moving the Data Warehouse into the Cloud	469
Dealing with Unstructured Data	470
<i>Summary</i>	470 • <i>Key Terms</i> 471 • <i>Review Questions</i> 471 •
<i>Problems and Exercises</i>	472 • <i>Field Exercises</i> 476 •
<i>References</i>	476 • <i>Further Reading</i> 477 •
<i>Web Resources</i>	477

Chapter 10 Big Data Technologies 478

Learning Objectives	478
Introduction	478
Moving Beyond Transactional and Data Warehousing Databases	480
Big Data	480
NoSQL	482
Classification of NoSQL DBMSs	484
KEY-VALUE STORES	484
DOCUMENT STORES	485
WIDE-COLUMN STORES	485
GRAPH-ORIENTED DATABASES	485
NoSQL Examples	485
REDIS	486
MONGODB	486
APACHE CASSANDRA	486
NEO4J	486
A NoSQL Example: MongoDB	486
DOCUMENTS	486
COLLECTIONS	488
RELATIONSHIPS	488
QUERYING MONGODB	488
Impact of NoSQL on Database Professionals	492
Hadoop	492
Components of Hadoop	492
THE HADOOP DISTRIBUTED FILE SYSTEM (HDFS)	493
MAPREDUCE	493
PIG	495
HIVE	495
HBASE	496
A Practical Introduction to Pig	496
LOADING DATA	496
TRANSFORMING DATA	497
A Practical Introduction to Hive	499
CREATING A TABLE	499
LOADING DATA INTO THE TABLE	499
PROCESSING THE DATA	500
Integrated Analytics and Data Science Platforms	502
HP HAVEN	502
TERADATA ASTER	502
IBM BIG DATA PLATFORM	503

Putting It All Together: Integrated Data Architecture 503

*Summary 505 • Key Terms 505 • Review Questions 505 •
Problems and Exercises 506 • References 506 •
Further Reading 507 • Web Resources 507*

Chapter 11 Analytics and Its Implications 508

Learning Objectives 508

Introduction 508

Analytics 509

Types of Analytics 509

Use of Descriptive Analytics 511

SQL OLAP QUERYING 512

OLAP TOOLS 514

DATA VISUALIZATION 516

BUSINESS PERFORMANCE MANAGEMENT AND DASHBOARDS 517

Use of Predictive Analytics 518

DATA MINING TOOLS 519

EXAMPLES OF PREDICTIVE ANALYTICS 520

Use of Prescriptive Analytics 521

Key User Tools for Analytics 522

ANALYTICAL AND OLAP FUNCTIONS 523

R 524

PYTHON 525

APACHE SPARK 526

Data Management Infrastructure for Analytics 526

Impact of Big Data and Analytics 529

Applications of Big Data and Analytics 529

BUSINESS 530

E-GOVERNMENT AND POLITICS 530

SCIENCE AND TECHNOLOGY 530

SMART HEALTH AND WELL-BEING 531

SECURITY AND PUBLIC SAFETY 531

Implications of Big Data Analytics and Decision Making 531

PERSONAL PRIVACY VERSUS COLLECTIVE BENEFITS 532

OWNERSHIP AND ACCESS 532

QUALITY AND REUSE OF DATA AND ALGORITHMS 532

TRANSPARENCY AND VALIDATION 532

CHANGING NATURE OF WORK 533

DEMANDS FOR WORKFORCE CAPABILITIES AND EDUCATION 533

Summary 533 • Key Terms 534 • Review Questions 534 •

Problems and Exercises 534 • References 535 •

Further Reading 536

Chapter 12 Data and Database Administration with Focus on Data Quality 537

Learning Objectives 537

Introduction 537

Overview of Data and Database Administration 539

Data Administration 539

Database Administration 540

TRADITIONAL DATABASE ADMINISTRATION	540
TRENDS IN DATABASE ADMINISTRATION	542
Evolving Data Administration Roles	544
The Open Source Movement and Database Management	545
Data Governance	546
Managing Data Quality	547
Characteristics of Quality Data	548
EXTERNAL DATA SOURCES	549
REDUNDANT DATA STORAGE AND INCONSISTENT METADATA	550
DATA ENTRY PROBLEMS	550
LACK OF ORGANIZATIONAL COMMITMENT	550
Data Quality Improvement	550
GET THE BUSINESS BUY-IN	550
CONDUCT A DATA QUALITY AUDIT	551
ESTABLISH A DATA STEWARDSHIP PROGRAM	552
IMPROVE DATA CAPTURE PROCESSES	552
APPLY MODERN DATA MANAGEMENT PRINCIPLES AND TECHNOLOGY	553
APPLY TQM PRINCIPLES AND PRACTICES	553
Summary of Data Quality	553
Data Availability	554
Costs of Downtime	554
Measures to Ensure Availability	555
HARDWARE FAILURES	555
LOSS OR CORRUPTION OF DATA	555
HUMAN ERROR	555
MAINTENANCE DOWNTIME	555
NETWORK-RELATED PROBLEMS	555
Master Data Management	555
<i>Summary</i>	<i>557</i>
<i>Key Terms</i>	<i>557</i>
<i>Review Questions</i>	<i>558</i>
<i>Problems and Exercises</i>	<i>558</i>
<i>Field Exercises</i>	<i>560</i>
<i>References</i>	<i>560</i>
<i>Further Reading</i>	<i>561</i>
<i>Web Resources</i>	<i>561</i>
<i>Glossary of Acronyms</i>	<i>563</i>
<i>Glossary of Terms</i>	<i>565</i>
<i>Index</i>	<i>573</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 13-9

Horizontal Partitioning 13-9

Vertical Partitioning 13-10

Combinations of Operations 13-11

Selecting the Right Data Distribution Strategy 13-12

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 13-19

Query Optimization 13-19

Evolution of Distributed DBMSs 13-22

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 13-27 •
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 2016 Notation A-1
 - ENTITIES A-5
 - RELATIONSHIPS A-5
 - CA ERwin Data Modeler 9.7 Notation A-5
 - ENTITIES A-5
 - RELATIONSHIPS A-5
 - SAP Sybase PowerDesigner 16.6 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 Resources](#) 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
 - [References](#) C-12

This page intentionally left blank

PREFACE

This text is designed for introductory courses 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 and MSIS 2016) all outline this type of database management course or the competencies a student completing the course is expected to have. Previous editions of this text have been used successfully for more than 35 years at both the undergraduate and graduate levels as well as in management and professional development programs.

WHAT'S NEW IN THIS EDITION?

This 13th 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 13th edition reflect the major trends in the information systems field and the skills required of modern information systems graduates. The most important changes are as follows:

- The book has been restructured in several important ways. Chapter 7 on databases in applications now also includes segments on transaction integrity, designing multi-user solutions, and application level security, bringing these important perspectives together with their context. The revised chapter on physical database design and database infrastructure (new Chapter 8) includes also coverage of database security, backup and recovery, cloud-based database solutions, and other essential database infrastructure topics. This new comprehensive structure on physical design and infrastructure is now placed after the SQL chapters. The new version of Chapter 9 integrates material on data warehousing and data integrity in a conceptually natural pairing. Recognizing the way in which analytics capabilities rely on all types of data management solutions, Chapter 11, on analytics and implications, is now separate from Chapter 10, on big data. Finally, Chapter 12 brings together data and database administration with data quality, emphasizing the essential connections between the three.
- The part structure of the book has been redesigned to be fully aligned with the new chapter structure.
- We have introduced a new overarching framework (Figure 1-5), which gives our readers a clearer overview of structure of the book and its core topic areas. The framework communicates clearly the increasing importance of informational systems (divided into Analytics–Data Warehousing and Analytics–Big Data) in addition to this book's traditional strength of transactional systems.
- Given the continued and still increasing interest in big data and analytics, we have continued to expand content in this area. The book has now separate chapters on big data technologies (Chapter 10) and analytics (Chapter 11). In addition to general coverage of NoSQL and Hadoop technologies, Chapter 10 provides also detailed examples of MongoDB, Pig, and Hive. Chapter 11 includes extended coverage of R, Python, and Apache Spark—all essential technologies for analytics professionals that allow a link between analytics and data management architectures.
- We emphasize the increasing importance of cloud-based database solutions, mobile technologies, and agile development throughout the book.
- Chapter 1 now better recognizes the broad range of enterprise level applications data management solutions enable and support, including enterprise systems, data warehouses, and data lakes.

- Chapter 7 on databases in applications now includes an extensive example demonstrating the use of Python in the context of database-driven applications.
- The instructor's manual will have more material to support the case Forondo Artist Management Excellence that was introduced in the 12th edition.

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.
- The entire book has been edited so that its language clearly reflects its focus on the readers as learners instead of authors as teachers
- End-of-chapter material (review questions, problems and exercises, and/or field exercises) in every chapter has been revised with new and modified questions and exercises.
- We continued to update the figures in several chapters 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 students have information on the latest database trends and expanded background details on important topics covered in the text.
- The book continues to be available through VitalSource, 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). Look for special icons on the opening page of these chapters to call attention to these videos, and go to www.pearsonglobaleditions.com 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 13th 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 13th edition, we continue this tradition by continuing to expand and improve coverage 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 13th 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 35 years as well as through our own teaching. Overall, the pedagogy of the book is sound, and we believe that the new framework that we introduced in Chapter 1 will further strengthen our students'

understanding of the big picture of data management. 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 5 and 6, 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 7 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 application development. 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 13th 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 to this chapter has been the introduction of a new integrated data management framework (Figure 1-5) and supporting text accompanying it. This framework recognizes the increasing importance of the *informational* systems in addition to the traditional focus of this book on *transactional* systems. 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 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, multi-tier, 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. The first chapter provides an introduction to the FAME case, which then continues through the book until Chapter 8.

PART II: DATABASE ANALYSIS AND LOGICAL DESIGN

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 E-R model: to unambiguously document the rules of the business that influence database design. 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 13th edition, we have provided six new problems and exercises; these new exercises present some more modern situations, such as Internet of Things applications for databases. A variety of other problems and exercises as well as review questions have been changed to emphasize important topics of the chapter. 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, with three new exercises and several building on or extending the new exercises from Chapter 2. 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.

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. New problems and exercises are included that draw upon the new problems and exercises from Chapters 2 and 3 for relational modeling and normalization. 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.

PART III: DATABASE IMPLEMENTATION AND USE

Chapter 5: Introduction to SQL

This chapter (Chapter 6 in 12th edition) presents a thorough introduction to the SQL used by most DBMSs (SQL:1999) and introduces the changes that are included in the latest standards (SQL: 2011 and SQL:2016). This edition adds coverage of the new features of SQL:2016, including row pattern recognition, JSON support, and extended analytical

capabilities. The new edition also clarifies coverage of SQL data types and, overall, makes it easier to move from relational design in Chapter 4 directly to database implementation without the material on physical database design (now in Chapter 8). The coverage of SQL is extensive and divided between this chapter and Chapter 6. This chapter includes examples of SQL code, using mostly SQL:1999 and SQL:2016 syntax, as well as some Oracle 12c and Microsoft SQL Server syntax. Some unique features of MySQL are mentioned. In this edition, coverage of views has been moved to Chapter 6. Chapter 5 explains the SQL commands needed to create and maintain a database and to program single-table queries. Five review questions and 13 problems and exercises have been added to the chapter or modified extensively. The chapter continues to use the Pine Valley Furniture Company case to illustrate a wide variety of practical queries and query results.

Chapter 6: Advanced SQL

This chapter (Chapter 7 in 12th edition) continues the description of SQL, with a careful explanation of multiple-table queries, transaction integrity, data dictionaries, dynamic and materialized views, triggers and stored procedures (the differences between them are now more clearly explained), and embedding SQL in other programming language programs. All forms of the OUTER JOIN command are covered. Standard SQL (with an updated focus on SQL:2016) is also used. The revised version of the chapter includes now thorough coverage of views and the purposes for which they are used, including their role in enabling security and privacy solutions. 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. The material of transaction integrity, has, however been moved to Chapter 7, where it most naturally belongs. The chapter continues to contain a clear explanation of subqueries and correlated subqueries, two of the most complex and powerful constructs in SQL. At the end, the chapter discusses material that is new to this chapter: data dictionary facilities (in practice, using SQL to understand the structure of the database) and recent extensions and enhancements to SQL. Chapter review material has been updated with 13 new problems and exercises and three new review questions.

Chapter 7: Databases in Applications

This chapter (Chapter 8 in 12th edition) provides a modern discussion of the concepts of client/server architecture and applications, middleware, and database access in contemporary database environments. The chapter has been structurally significantly modified to provide additional clarity, including the integration of material on a two-tiered architecture into the section on three-tiered architecture. In addition to a revised example of writing a Java web application, there is an entire new section—including an extensive and detailed example—on writing Web applications with Python, a widely used general purpose programming language that has become very popular in analytics. Sections on transaction integrity, concurrent access, and application level data security have been revised and moved to this chapter to provide additional conceptual clarity. Material on cloud computing has been moved to Chapter 8 on database infrastructure. Review questions and problems and exercises have been updated.

Chapter 8: Physical Database Design and Database Infrastructure

This chapter (Chapter 5 in the 12th edition) 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. In addition, several new topics on database infrastructure have been integrated into this chapter to improve the structural clarity of the book, including data dictionaries and repositories, general database software security features, and database backup and recovery. A revised and extended section on cloud-based database infrastructure completes the chapter. Overall, the chapter emphasizes ways to

improve database performance, with references to specific techniques available in Oracle and other DBMSs to achieve this goal. The discussion of indexes includes descriptions of the types of indexes 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. Review questions and problems and exercises have been updated and extended based on the new structure and content of the chapter.

PART IV: ADVANCED DATABASE TOPICS

Chapter 9: Data Warehousing and Data Integration

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. The most important change of this chapter is the integration of material on data integration (formerly in Chapter 10 in the 12th edition) into it. This change strengthens the readers' ability to understand the essential role of data integration in data warehousing (particularly in ETL and other aspects of data preparation), and it clarifies the structure of the book. Topics covered in this chapter include alternative data warehouse architectures and the dimensional data model (or star schema) for data warehouses. In this edition, additional attention is given to cloud-based implementation of data warehouses. Throughout the chapter, several details have been updated to ensure technical correctness. Operational data store and independent, dependent, and logical data marts are defined. The chapter includes multiple new and revised review questions and problems and exercises.

Chapter 10: Big Data Technologies

This chapter incorporates big data infrastructure material from Chapter 11 in the 12th edition, significantly expanding it and making it more directly applicable with substantial detailed descriptive examples of MongoDB (the most popular NoSQL database) and Pig (scripting language and task automation environment for Hadoop) and Hive (an SQL-like declarative language for querying data stored in Hadoop). This new version of the material gives the students a much more practical, hands-on sense of the purposes for which these well-known tools can be used and how they can serve the goals of big data management. The chapter also includes several new problems and exercises based on these environments. Overall, the chapter helps the readers understand how big data technologies have expanded the possibilities for analytics-driven innovation through advanced informational systems that are pushing boundaries further in terms of volume, velocity, and variety of data while paying continuous attention to value and veracity of big data.

Chapter 11: Analytics and its Implications

Chapter 11 offers integrated coverage of analytics, including descriptive, predictive, and prescriptive analytics. It is based on material on analytics in the big data and analytics chapter in the 12th edition, expanding it with comprehensive new sections on R, Python, and Apache Spark and bringing in material on analytical functions in SQL. 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 (as indicated in the new framework in Chapter 1). The chapter also covers approaches and technologies used by analytics professionals, such as on-line analytical processing, data visualization, business performance management and dashboards, data mining, and text mining. Finally, the chapter integrates the coverage of big data and analytics technologies to the individual, organizational, and societal implications of these capabilities. Review questions on the new material have been added.

Chapter 12: Data and Database Administration with Focus on Data Quality

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 a renewed and strengthened emphasis on data quality. The chapter both discusses essential characteristics of high-quality data and the mechanisms that organizations need to put in place to enable data quality improvement. Data governance, data availability, and master data management are also covered. The chapter continues to emphasize the critical importance of data and database management in managing data as a corporate asset.

Chapter 13: Distributed Databases

This chapter—available on the book’s Web site—reviews the role, technologies, and unique database design opportunities of distributed databases. The objectives and trade-offs for distributed databases, data replication alternatives, factors in selecting a data distribution strategy, and distributed database vendors and products are covered. This chapter provides thorough coverage of database concurrency access controls. Many reviewers have indicated that they are seldom able to cover this chapter in an introductory course, but having the material available is critical for advanced students or special topics.

Chapter 14: Object-Oriented Data Modeling

This chapter presents an introduction to object-oriented modeling using Object Management Group’s Unified Modeling Language (UML). This chapter has been carefully reviewed to ensure consistency with the latest UML notation and best industry practices. UML provides an industry-standard notation for representing classes and objects. The chapter continues to emphasize basic object-oriented concepts, such as inheritance, encapsulation, composition, and polymorphism. As with Chapter 13, Chapter 14 is available on the textbook’s Web site.

APPENDICES

In the 13th edition three appendices are available on the book’s Web site and are intended for those who wish to explore certain topics in greater depth.

Appendix A: Data Modeling Tools and Notation

This appendix addresses a need raised by many readers—how to translate the E-R notation in the text into the form used by the CASE tool or the DBMS used in class. Specifically, this appendix compares the notations of CA ERwin Data Modeler r9.7, Oracle SQL Data Modeler 4.2, SAP Sybase PowerDesigner 16.6, and Microsoft Visio Professional 2016. Tables and illustrations show the notations used for the same constructs in each of these popular software packages.

Appendix B: Advanced Normal Forms

This appendix presents a description (with examples) of Boyce-Codd and fourth normal forms, including an example of BCNF to show how to handle overlapping candidate keys. Other normal forms are briefly introduced. The Web Resources section includes a reference for information on many advanced normal form topics.

Appendix C: Data Structures

This appendix describes several data structures that often underlie database implementations. Topics include the use of pointers, stacks, queues, sorted lists, inverted lists, and trees.

PEDAGOGY

A number of additions and improvements have been made to end-of-chapter materials to provide a wider and richer range of choices for the user. The most important of these improvements are the following:

1. **Review Questions** Questions have been updated to support new and enhanced chapter material.
2. **Problems and Exercises** This section has been reviewed in every chapter, and many chapters contain new problems and exercises to support updated chapter material. Of special interest are questions in many chapters that give students opportunities to use the data sets provided for the text. Problems and exercises are presented in roughly increasing order of difficulty, which should help instructors and students find exercises appropriate for what they want to accomplish.
3. **Field Exercises** This section provides a set of “hands-on” mini-cases that can be assigned to individual students or to small teams of students. Field exercises range from directed field trips to Internet searches and other types of research exercises.
4. **Case** The 13th edition of this book includes the same mini-case that was introduced in the 12th edition: Forondo Artist Management Excellence Inc. (FAME). In the first three chapters, the case begins with a description provided in the “voice” of one or more stakeholders, revealing a new dimension of requirements to the reader. Each chapter has project assignments intended to provide guidance on the types of deliverables instructors could expect from students, some of which tie together issues and activities across chapters. These project assignments can be completed by individual students or by small project teams. This case provides an excellent means for students to gain hands-on experience with the concepts and tools they have studied. The instructor’s manual will include new materials to support the use of the case.
5. **Web Resources** Each chapter contains a list of updated and validated URLs for Web sites that contain information that supplements the chapter. These Web sites cover online publication archives, vendors, electronic publications, industry standards organizations, and many other sources. These sites allow students and instructors to find updated product information, innovations that have appeared since the printing of the book, background information to explore topics in greater depth, and resources for writing research papers.

We continue to provide several pedagogical features that help make the 13th edition widely accessible to instructors and students. These features include the following:

1. **Learning objectives** appear at the beginning of each chapter, as a preview of the major concepts and skills students will learn from that chapter. The learning objectives—carefully updated to be aligned with the new chapter structure—also provide a great study review aid for students as they prepare for assignments and examinations.
2. **Chapter introductions and summaries** both encapsulate the main concepts of each chapter and link material to related chapters, providing students with a comprehensive conceptual framework for the course.
3. **The chapter review** includes the Review Questions, Problems and Exercises, and Field Exercises discussed earlier and also contains a Key Terms list to test the student’s grasp of important concepts, basic facts, and significant issues.
4. **A running glossary** defines key terms in the page margins as they are discussed in the text. These terms are also defined at the end of the text, in the Glossary of Terms. Also included is the end-of-book Glossary of Acronyms for abbreviations commonly used in database management.

ORGANIZATION

We encourage instructors to customize their use of this book to meet the needs of both their curriculum and student career paths. The modular nature of the text, its broad coverage, its extensive illustrations, and its inclusion of advanced topics and emerging issues make customization easy. The many references to current publications and Web sites