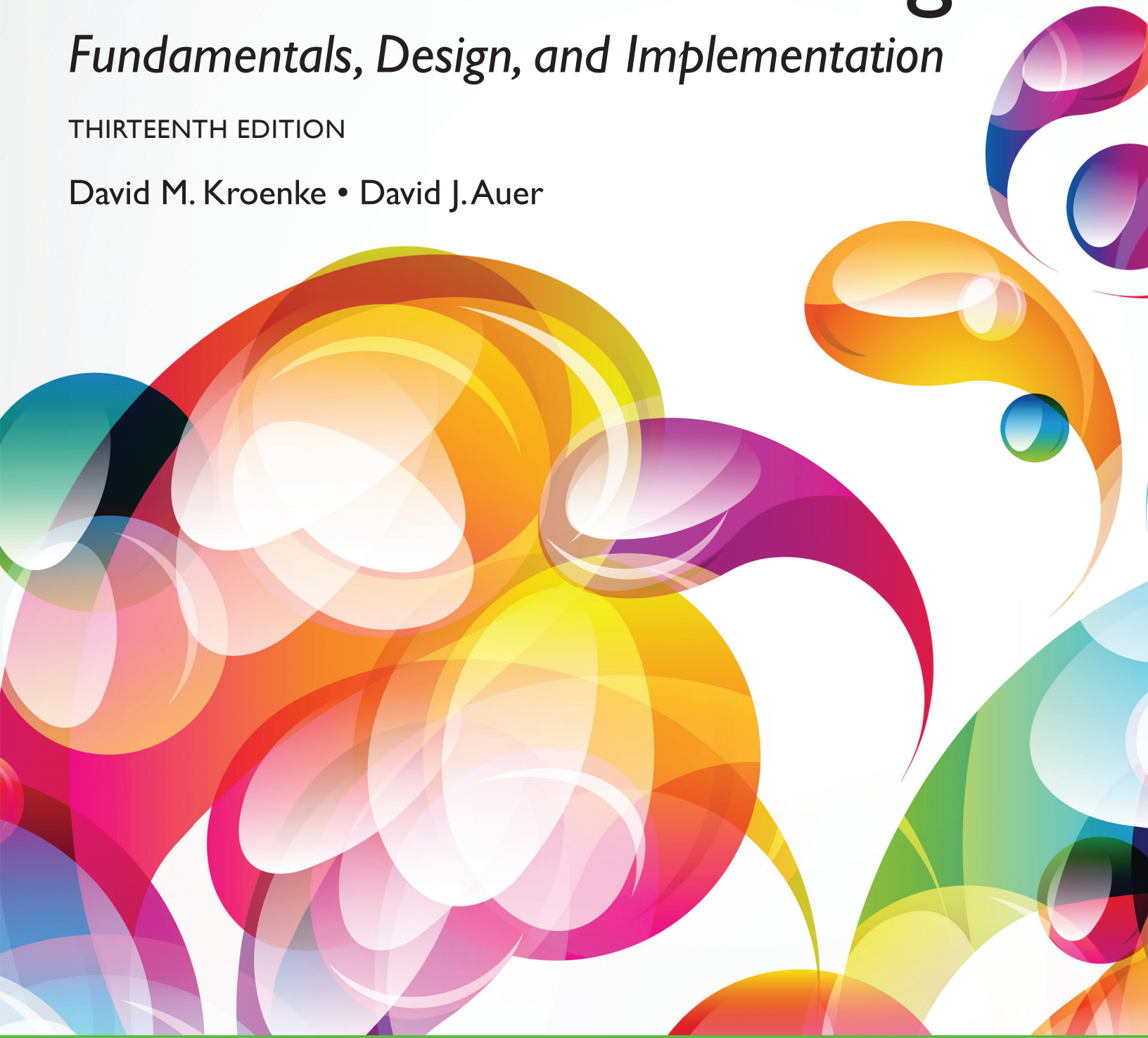
INTERNATIONAL EDITION

Database Processing

Fundamentals, Design, and Implementation

THIRTEENTH EDITION

David M. Kroenke • David J. Auer



ALWAYS LEARNING

PEARSON

Database Processing

Fundamentals, Design, and Implementation

EDITION 13

Database Processing

Fundamentals, Design, and Implementation

EDITION 13

David M. Kroenke

David J. Auer

Western Washington University

International Edition contributions by

Arup Kumar Bhattacharjee

RCC Institute of Information Technology, Kolkata

Soumen Mukherjee

RCC Institute of Information Technology, Kolkata

PEARSON

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

Editor in Chief: Stephanie Wall
Executive Editor: Bob Horan
Program Management Lead: Ashley Santora
Program Manager: Kelly Loftus
Editorial Assistant: Kaylee Rotella
Director of Marketing: Maggie Moylan
Executive Marketing Manager: Anne Fahlgren
Marketing Assistant: Gianna Sandri
Project Management Lead: Judy Leale
Publishing Operations Director, International Edition: Angshuman Chakraborty
Manager, Publishing Operations, International Edition: Shokhi Shah Khandelwal
Associate Print & Media Editor, International Edition: Anuprova Dey Chowdhuri
Acquisitions Editor, International Edition: Shivangi Ramachandran
Publishing Administrator, International Edition: Hema Mehta
Project Editor, International Edition: Karthik Subramanian
Senior Manufacturing Controller, Production, International Edition: Trudy Kimber
Production Project Manager: Jane Bonnell
Operations Specialist: Michelle Klein
Senior Art Director: Janet Slowik
Interior Designer: Karen Quigley
Cover Designer: Jodi Notowitz
Cover Image: Itana/Shutterstock.com
Media Project Manager, Editorial: Denise Vaughn
Media Project Manager, Production: Lisa Rinaldi

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.pearsoninternationaleditions.com

© Pearson Education Limited 2014

The rights of David M. Kroenke and David J. Auer to be identified as 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 Database Processing: Fundamentals, Design, and Implementation, 13th Edition, ISBN 978-0-13-305835-2, *by David M. Kroenke and David J. Auer, published by Pearson Education © 2014.*

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.

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.

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.

ISBN 10: 1-292-00486-X
ISBN 13: 978-1-292-00486-0

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
14 13 12 11 10

Typeset in Kepler Std-Light by Integra

Printed and bound by Courier Kendallville in The United States of America

Brief Contents

Part 1 • Getting Started 33

Chapter 1 Introduction 34
Chapter 2 Introduction to Structured Query Language 65

Part 2 • Database Design 149

Chapter 3 The Relational Model and Normalization 150
Chapter 4 Database Design Using Normalization 192
Chapter 5 Data Modeling with the Entity-Relationship Model 211
Chapter 6 Transforming Data Models into Database Designs 262

Part 3 • Database Implementation 309

Chapter 7 SQL for Database Construction and Application Processing 310
Chapter 8 Database Redesign 398

Part 4 • Multiuser Database Processing 425

Chapter 9 Managing Multiuser Databases 426
Chapter 10 Managing Databases with SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6 460

Online Chapter: See Page 465 for Instructions
Chapter 10A Managing Databases with SQL Server 2012

Online Chapter: See Page 465 for Instructions
Chapter 10B Managing Databases with Oracle Database 11*g* Release 2

Online Chapter: See Page 465 for Instructions
Chapter 10C Managing Databases with MySQL 5.6

Part 5 • Database Access Standards 467

Chapter 11 The Web Server Environment 468
Chapter 12 Big Data, Data Warehouses, and Business Intelligence Systems 566

Online Appendices: See Page 612 for Instructions
Appendix A Getting Started with Microsoft Access 2013
Appendix B Getting Started with Systems Analysis and Design
Appendix C E-R Diagrams and the IDEF1X Standard
Appendix D E-R Diagrams and the UML Standard
Appendix E Getting Started with MySQL Workbench Data Modeling Tools
Appendix F Getting Started with Microsoft Visio 2013
Appendix G Data Structures for Database Processing
Appendix H The Semantic Object Model
Appendix I Getting Started with Web Servers, PHP, and the Eclipse PDT
Appendix J Business Intelligence Systems

Contents

Preface 19

Part 1 • Getting Started 33

Chapter 1: Introduction 34

Chapter Objectives 34
The Characteristics of Databases 35
A Note on Naming Conventions 36 • A Database Has Data and Relationships 36 • Databases Create Information 37
Database Examples 38
Single-User Database Applications 38 • Multiuser Database Applications 38 • E-Commerce Database Applications 39 • Reporting and Data Mining Database Applications 39
The Components of a Database System 40
Database Applications and SQL 41 • The DBMS 43 • The Database 44
Personal Versus Enterprise-Class Database Systems 46
What Is Microsoft Access? 46 • What Is an Enterprise-Class Database System? 47
Database Design 49
Database Design from Existing Data 49 • Database Design for New Systems Development 50 • Database Redesign 51
What You Need to Learn 52
A Brief History of Database Processing 53
The Early Years 53 • The Emergence and Dominance of the Relational Model 55 • Post-Relational Developments 56

Summary 58 • Key Terms 59 • Review Questions 59 • Project Questions 61

Chapter 2: Introduction to Structured Query Language 65

Chapter Objectives 65
Components of a Data Warehouse 66
Cape Codd Outdoor Sports 67
The Extracted Retail Sales Data 67 • RETAIL_ORDER Data 68 • ORDER_ITEM Data 69 • SKU_DATA Table 70 • The Complete Cape Codd Data Extract Schema 70 • Data Extracts Are Common 71
SQL Background 71
The SQL SELECT/FROM/WHERE Framework 72
Reading Specified Columns from a Single Table 72 • Specifying Column Order in SQL Queries from a Single Table 73 • Reading Specified Rows from a Single Table 75 • Reading Specified Columns and Rows from a Single Table 76
Submitting SQL Statements to the DBMS 77
Using SQL in Microsoft Access 2013 77 • Using SQL in Microsoft SQL Server 2012 82 • Using SQL in Oracle Database 11g Release 2 85 • Using SQL in Oracle MySQL 5.6 87

SQL Enhancements for Querying a Single Table 90
Sorting the SQL Query Results 90 • SQL WHERE Clause Options 92 • Combing the SQL WHERE Clause and the SQL ORDER BY Clause 97
Performing Calculations in SQL Queries 97
Using SQL Built-in Functions 97 • SQL Expressions in SQL SELECT Statements 100
Grouping in SQL SELECT Statements 102
Looking for Patterns in NASDAQ Trading 106
Investigating the Characteristics of the Data 106 • Searching for Patterns in Trading by Day of Week 107
Querying Two or More Tables with SQL 109
Querying Multiple Tables with Subqueries 109 • Querying Multiple Tables with Joins 112 • Comparing Subqueries and Joins 117 • The SQL JOIN ON Syntax 117 • Outer Joins 119

Summary 123 • Key Terms 123 • Review Questions 124 • Project Questions 129 • Case Questions 133 • The Queen Anne Curiosity Shop 138 • Morgan Importing 145

Part 2 • Database Design 149

Chapter 3: The Relational Model and Normalization 150

Chapter Objectives 150
Relational Model Terminology 152
Relations 152 • Characteristics of Relations 153 • Alternative Terminology 155 • Functional Dependencies 156 • Finding Functional Dependencies 158 • Keys 161
Normal Forms 163
Modification Anomalies 163 • A Short History of Normal Forms 164 • Normalization Categories 165 • From First Normal Form to Boyce-Codd Normal Form Step by Step 166 • Eliminating Anomalies from Functional Dependencies with BCNF 169 • Eliminating Anomalies from Multivalued Dependencies 179 • Fifth Normal Form 183 • Domain/Key Normal Form 183

Summary 183 • Key Terms 184 • Review Questions 185 • Project Questions 187 • Case Questions 188 • The Queen Anne Curiosity Shop 189 • Morgan Importing 191

Chapter 4: Database Design Using Normalization 192

Chapter Objectives 192
Assess Table Structure 193
Designing Updatable Databases 194
Advantages and Disadvantages of Normalization 194 • Functional Dependencies 195 • Normalizing with SQL 195 • Choosing Not to Use BCNF 196 • Multivalued Dependencies 197
Designing Read-Only Databases 197
Denormalization 198 • Customized Duplicated Tables 199
Common Design Problems 199
The Multivalue, Multicolumn Problem 200 • Inconsistent Values 202 • Missing Values 203 • The General-Purpose Remarks Column 204

Summary 205 • Key Terms 205 • Review Questions 206 • Project Questions 208 • Case Questions 208 • The Queen Anne Curiosity Shop 209 • Morgan Importing 210

Chapter 5: Data Modeling with the Entity-Relationship Model 211

Chapter Objectives 211
The Purpose of a Data Model 212
The Entity-Relationship Model 212
Entities 213 • Attributes 213 • Identifiers 213 • Relationships 214 • Maximum Cardinality 216 • Minimum Cardinality 217 • Entity-Relationship Diagrams and Their Versions 218 • Variations of the E-R Model 218 • E-R Diagrams Using the IE Crow's Foot Model 219 • Strong Entities and Weak Entities 221 • ID-Dependent Entities 221 • Non-ID-Dependent Weak Entities 222 • The Ambiguity of the Weak Entity 223 • Subtype Entities 223
Patterns in Forms, Reports, and E-R Models 225
Strong Entity Patterns 226 • ID-Dependent Relationships 228 • Mixed Identifying and Nonidentifying Patterns 235 • The For-Use-By Pattern 237 • Recursive Patterns 239
The Data Modeling Process 241
The College Report 242 • The Department Report 243 • The Department/Major Report 245 • The Student Acceptance Letter 246

Summary 248 • Key Terms 249 • Review Questions 249 • Project Questions 251 • Case Questions 257 • The Queen Anne Curiosity Shop 260 • Morgan Importing 261

Chapter 6: Transforming Data Models into Database Designs 262

Chapter Objectives 262
The Purpose of a Database Design 263
Create a Table for Each Entity 263
Selecting the Primary Key 264 • Specifying Candidate (Alternate) Keys 265 • Specify Column Properties 266 • Verify Normalization 268
Create Relationships 268
Relationships Between Strong Entities 269 • Relationships Using ID-Dependent Entities 272 • Relationships with a Weak Non-ID-Dependent Entity 277 • Relationships in Mixed Entity Designs 277 • Relationships Between Supertype and Subtype Entities 279 • Recursive Relationships 279 • Representing Ternary and Higher-Order Relationships 281 • Relational Representation of the Highline University Data Model 284
Design for Minimum Cardinality 285
Actions When the Parent Is Required 287 • Actions When the Child Is Required 288 • Implementing Actions for M-O Relationships 288 • Implementing Actions for O-M Relationships 289 • Implementing Actions for M-M Relationships 290 • Designing Special Case M-M Relationships 290 • Documenting the Minimum Cardinality Design 291 • An Additional Complication 293 • Summary of Minimum Cardinality Design 293
The View Ridge Gallery Database 293
Summary of Requirements 293 • The View Ridge Gallery Data Model 294 • Database Design with Data Keys 295 • Minimum Cardinality Enforcement for Required Parents 296 • Minimum Cardinality Enforcement for the Required Child 297 • Column Properties for the View Ridge Gallery Database Design Tables 300

Summary 301 • Key Terms 302 • Review Questions 302 • Project Questions 304 • Case Questions 305 • The Queen Anne Curiosity Shop 307 • Morgan Importing 307

Part 3 • Database Implementation 309

Chapter 7: SQL for Database Construction and Application Processing 310

Chapter Objectives 310
The Importance of Working with an Installed DBMS Product 311
The View Ridge Gallery Database 311
SQL DDL and DML 312
Managing Table Structure with SQL DDL 313
Creating the View Ridge Gallery Database 313 • Using SQL Scripts 313 • Using the SQL CREATE TABLE Statement 314 • Variations in SQL Data Types 315 • Creating the ARTIST Table 315 • Creating the WORK Table and the 1:N ARTIST-to-WORK Relationship 322 • Implementing Required Parent Rows 324 • Implementing 1:1 Relationships 324 • Casual Relationships 324 • Creating Default Values and Data Constraints with SQL 324 • Creating the View Ridge Gallery Database Tables 326 • The SQL ALTER TABLE Statement 330 • The SQL DROP TABLE Statement 330 • The SQL TRUNCATE TABLE Statement 331 • The SQL CREATE INDEX Statement 331
SQL DML Statements 332
The SQL INSERT Statement 332 • Populating the View Ridge Gallery Database Tables 333 • The SQL UPDATE Statement 339 • The SQL MERGE Statement 340 • The SQL DELETE Statement 341
Using SQL Views 341
Using SQL Views to Hide Columns and Rows 343 • Using SQL Views to Display Results of Computed Columns 345 • Using SQL Views to Hide Complicated SQL Syntax 346 • Layering Built-in Functions 346 • Using SQL Views for Isolation, Multiple Permissions, and Multiple Triggers 349 • Updating SQL Views 350
Embedding SQL in Program Code 351
SQL/Persistent Stored Modules (SQL/PSM) 352 • Using SQL User-Defined Functions 352 • Using SQL Triggers 355 • Using Stored Procedures 362

Summary 365 • Key Terms 366 • Review Questions 367 • Project Questions 371 • Case Questions 375 • The Queen Anne Curiosity Shop 385 • Morgan Importing 391

Chapter 8: Database Redesign 398

Chapter Objectives 398
The Need for Database Redesign 399
SQL Statements for Checking Functional Dependencies 399
What Is a Correlated Subquery? 400 • EXISTS and NOT EXISTS 403
How Do I Analyze an Existing Database? 405
Reverse Engineering 405 • Dependency Graphs 406 • Database Backup and Test Databases 407
Changing Table Names and Table Columns 408
Changing Table Names 408 • Adding and Dropping Columns 410 • Changing a Column Data Type or Column Constraints 411 • Adding and Dropping Constraints 411
Changing Relationship Cardinalities 411
Changing Minimum Cardinalities 412 • Changing Maximum Cardinalities 413 • Reducing Cardinalities (with Data Loss) 415
Adding and Deleting Tables and Relationships 416
Forward Engineering 416

Summary 417 • Key Terms 418 • Review Questions 418 • Project Questions 420 • Case Questions 421 • The Queen Anne Curiosity Shop 422 • Morgan Importing 423

Part 4 • Multiuser Database Processing 425

Chapter 9: Managing Multiuser Databases 426

Chapter Objectives 426
The Importance of Working with an Installed DBMS Product 427
Database Administration 428
Managing the Database Structure 428
Concurrency Control 429
The Need for Atomic Transactions 430 • Resource Locking 433 • Optimistic Versus Pessimistic Locking 435
SQL Transaction Control Language and Declaring Lock Characteristics 436
Implicit and Explicit COMMIT TRANSACTION 438 • Consistent Transactions 438 • Transaction Isolation Level 439 SQL Cursors 440
Database Security 442
Processing Rights and Responsibilities 442 • DBMS Security 443 • DBMS Security Guidelines 444 • Application Security 446 • The SQL Injection Attack 447
Database Backup and Recovery 447
Recovery via Reprocessing 448 • Recovery via Rollback/Rollforward 448
Managing the DBMS 450
Maintaining the Data Repository 451

Summary 452 • Key Terms 453 • Review Questions 454 • Project Questions 456 • Case Questions 457 • The Queen Anne Curiosity Shop 458 • Morgan Importing 459

Chapter 10: Managing Databases with SQL Server 2012, Oracle 11*g* Release 2, and MySQL 5.6 460

Chapter Objectives 460
Installing the DBMS 461
Using the DBMS Database Administration and Database Development Utilities 462
Creating a Database 462
Creating and Running SQL Scripts 463
Reviewing the Database Structure in the DBMS GUI Utility 463
Creating and Populating the View Ridge Gallery Database Tables 463
Creating SQL Views for the View Ridge Gallery Database 464
Database Application Logic and SQL/Persistent Stored Modules (SQL/PSM) 464
DBMS Concurrency Control 464
DBMS Security 464
DBMS Backup and Recovery 464
Other DBMS Topics Not Discussed 465
Choose Your DBMS Product(s)! 465

Summary 465 • Key Terms 466 • Project Questions 466

ONLINE CHAPTER: SEE PAGE 465 FOR INSTRUCTIONS

Chapter 10A: Managing Databases with SQL Server 2012

Chapter Objectives
Installing SQL Server 2012
The Microsoft SQL Server 2012 Management Studio
Creating an SQL Server 2012 Database
SQL Server 2012 Utilities
SQL CMD and Microsoft PowerShell • Microsoft SQL CLR • SQL Server 2012 GUI Displays

SQL Server 2012 SQL Statements and SQL Scripts
Creating and Populating the View Ridge Gallery Database Tables
Creating the View Ridge Gallery Database Table Structure • *Reviewing Database Structures in the SQL Server GUI Display* • *Indexes* • *Populating the VRG Tables with Data* • *Creating Views*
SQL Server Application Logic
Transact-SQL • *User-Defined Functions* • *Stored Procedures* • *Triggers*
Concurrency Control
Transaction Isolation Level • *Cursor Concurrency* • *Locking Hints*
SQL Server 2012 Security
SQL Server Database Security Settings
SQL Server 2012 Backup and Recovery
Backing Up a Database • *SQL Server Recovery Models* • *Restoring a Database* • *Database Maintenance Plans*
Topics Not Discussed in This Chapter

Summary • Key Terms • Review Questions • Project Questions • Case Questions • The Queen Anne Curiosity Shop • Morgan Importing

ONLINE CHAPTER: SEE PAGE 465 FOR INSTRUCTIONS

Chapter 10B: Managing Databases with Oracle Database 11*g* Release 2

Chapter Objectives
Installing Oracle Database 11*g* Release 2
Installing a Loopback Adapter • *Oracle Database and Java* • *Oracle Database 11g Release 2 Documentation* • *The Oracle Universal Installer (OUI)*
Oracle Database 11*g* Release 2 Administration and Development Tools
The Oracle Database 11g Release 2 Configuration Assistant • *The Oracle Enterprise Manager 11g Database Control*
Oracle Database Tablespaces
Oracle Database Security
User Privileges • *Creating a User Account* • *Creating a Role*
Oracle Application Development Tools
*Oracle SQL*Plus* • *Oracle SQL Developer* • *Oracle Database Schemas*
Oracle Database 11*g* Release 2 SQL Statements and SQL Scripts
Creating and Populating the View Ridge Gallery Database Tables
Creating the View Ridge Gallery Database Table Structure • *Transaction COMMIT in Oracle Database* • *Reviewing Database Structures in the SQL Developer GUI Display* • *Indexes* • *Populating the VRG Tables* • *Creating Views*
Oracle Database Application Logic
Oracle Database PL/SQL • *User-Defined Functions* • *Stored Procedures* • *Triggers*
Oracle Database Concurrency Control
Read-Committed Transaction Isolation Level • *Serializable Transaction Isolation Level* • *Read-Only Transaction Isolation* • *Additional Locking Comments*
Oracle Database Backup and Recovery
Oracle Recovery Facilities • *Types of Failure*
Topics Not Discussed in This Chapter

Summary • Key Terms • Review Questions • Project Questions • Case Questions • The Queen Anne Curiosity Shop • Morgan Importing

ONLINE CHAPTER: SEE PAGE 465 FOR INSTRUCTIONS

Chapter 10C: Managing Databases with MySQL 5.6

Chapter Objectives
The MySQL 5.6 DBMS
Installing and Updating MySQL
Configuring MySQL • MySQL Storage Engines
The MySQL GUI Utilities
Creating a Workspace for the MySQL Workbench Files
Creating and Using a MySQL Database
Creating a Database in MySQL • Setting the Active Database in MySQL
MySQL Utilities
MySQL Command-Line Client • MySQL GUI Displays • MySQL SQL Statements and SQL Scripts
Creating and Populating the View Ridge Gallery Database Tables
Creating the View Ridge Gallery Database Table Structure • Reviewing Database Structures in the MySQL GUI Display • Indexes • Populating the VRG Tables with Data • Transaction COMMIT in MySQL • Creating Views
MySQL Application Logic
MySQL Procedural Statements • User-Defined Functions • Stored Procedures • Triggers • A Last Word on MySQL Stored Procedures and Triggers
Concurrency Control
MySQL 5.6 Security
MySQL Database Security Settings
MySQL 5.6 DBMS Backup and Recovery
Backing Up a MySQL Database • Restoring a MySQL Database
Topics Not Discussed in This Chapter

Summary • Key Terms • Review Questions • Project Questions • Case Questions • The Queen Anne Curiosity Shop • Morgan Importing

Part 5 • Database Access Standards 467

Chapter 11: The Web Server Environment 468

Chapter Objectives 468
A Web Database Application for the View Ridge Gallery 469
The Web Database Processing Environment 469
The ODBC Standard 471
ODBC Architecture 472 • Conformance Levels 473 • Creating an ODBC Data Source Name 474
The Microsoft .NET Framework and ADO.NET 480
OLE DB 481 • ADO and ADO.NET 485 • The ADO.NET Object Model 486
The Java Platform 489
JDBC 489 • Java Server Pages (JSP) and Servlets 492 • Apache Tomcat 492
Web Database Processing with PHP 493
Web Database Processing with PHP and Eclipse 494 • Getting Started with HTML Web Pages 496 • The index.html Web Page 497 • Creating the index.html Web Page 497 • Using PHP 499 • Challenges for Web Database Processing 505
Web Page Examples with PHP 506
Example 1: Updating a Table 507 • Example 2: Using PHP Data Objects (PDO) 511 • Example 3: Invoking a Stored Procedure 513
The Importance of XML 518
XML as a Markup Language 519
XML Document Type Declarations 519 • Materializing XML Documents with XSLT 520

XML Schema 524
XML Schema Validation 525 • Elements and Attributes 526 • Flat Versus Structured Schemas 527 • Global Elements 531
Creating XML Documents from Database Data 533
Using the SQL SELECT…FOR XML Statement 533 • Multitable SELECT with FOR XML 538 • An XML Schema for All CUSTOMER Purchases 542 • A Schema with Two Multivalued Paths 546
Why Is XML Important? 546
Additional XML Standards 551

Summary 554 • Key Terms 556 • Review Questions 557 • Project Questions 561 • Case Questions 562 • The Queen Anne Curiosity Shop 563 • Morgan Importing 564

Chapter 12: Big Data, Data Warehouses, and Business Intelligence Systems 566

Chapter Objectives 566
Business Intelligence Systems 568
The Relationship Between Operational and BI Systems 568
Reporting Systems and Data Mining Applications 569
Reporting Systems 569 • Data Mining Applications 569
Data Warehouses and Data Marts 570
Components of a Data Warehouse 570 • Data Warehouses Versus Data Marts 572 • Dimensional Databases 574
Reporting Systems 582
RFM Analysis 582 • OLAP 583
Data Mining 592
Distributed Database Processing 593
Types of Distributed Databases 593 • Challenges of Distributed Databases 594
Object-Relational Databases 595
Virtualization 596
Cloud Computing 598
Big Data and the Not Only SQL Movement 600
Structured Storage 600 • MapReduce 601 • Hadoop 602

Summary 603 • Key Terms 604 • Review Questions 605 • Project Questions 607 • Case Questions 608 • The Queen Anne Curiosity Shop 609 • Morgan Importing 610

Appendices

ONLINE APPENDICES: SEE PAGE 612 FOR INSTRUCTIONS

Appendix A: Getting Started with Microsoft Access 2013

Chapter Objectives
What Is the Purpose of This Appendix?
Why Should I Learn to Use Microsoft Access 2013?
What Will This Appendix Teach Me?
What Is a Table Key?
Relationships Among Tables
Creating a Microsoft Access Database
The Microsoft Office Fluent User Interface
The Ribbon and Command Tabs • Contextual Command Tabs • Modifying the Quick Access Toolbar • Database Objects and the Navigation Pane
Closing a Database and Exiting Microsoft Access

Opening an Existing Microsoft Access Database
Creating Microsoft Access Database Tables
Inserting Data into Tables—The Datasheet View
Modifying and Deleting Data in Tables in the Datasheet View
Creating Relationships Between Tables
Working with Microsoft Access Queries
Microsoft Access Forms and Reports
Closing a Database and Exiting Microsoft Access 2013

Key Terms • Review Questions

Appendix B: Getting Started with Systems Analysis and Design

Chapter Objectives
What Is the Purpose of This Appendix?
What Is Information?
What Is an Information System?
What Is a Competitive Strategy?
How Does a Company Organize Itself Based on Its Competitive Strategy?
What Is a Business Process?
How Do Information Systems Support Business Processes?
Do Information Systems Include Processes?
Do We Have to Understand Business Processes in Order to Create Information Systems?
What Is Systems Analysis and Design?
What Are the Steps in the SDLC?
The Ribbon and Command Tabs • The System Definition Step • The Requirements Analysis Step • The Component Design Step • The Implementation Step • The System Maintenance Step
What SDLC Details Do We Need to Know?
What Is Business Process Modeling Notation?
What Is Project Scope?
How Do I Gather Data and Information About System Requirements?
How Do Use Cases Provide Data and Information About System Requirements?
The Highline University Database
The College Report • The Department Report • The Department/Major Report • The Student Acceptance Letter
What Are Business Rules?
What Is a User Requirements Document (URD)?
What Is a Statement of Work (SOW)?

Key Terms • Review Questions • Project Questions

Appendix C: E-R Diagrams and the IDEF1X Standard

Chapter Objectives
IDEF1X Entities
IDEF1X Relationships
Nonidentifying Connection Relationships • Identifying Connection Relationships • Nonspecific Relationships • Categorization Relationships
Domains
Domains Reduce Ambiguity • Domains Are Useful • Base Domains and Typed Domains

Key Terms • Review Questions

Appendix D: E-R Diagrams and the UML Standard

Chapter Objectives
UML Entities and Relationships
Representation of Weak Entities
Representation of Subtypes
OOP Constructs Introduced by UML
The Role of UML in Database Processing Today

Key Terms • Review Questions

Appendix E: Getting Started with MySQL Workbench Data Modeling Tools

Chapter Objectives
What Is the Purpose of This Appendix?
Why Should I Learn to Use the MySQL Workbench Data Modeling Tools?
What Will This Appendix Teach Me?
What Won't This Appendix Teach Me?
How Do I Start the MySQL Workbench?
How Do I Create a Workspace for the MySQL Workbench Files?
How Do I Install the MySQL Connector/ODBC?
How Do I Create Database Designs in the MySQL Workbench?
How Do I Create a Database Model and E-R Diagram in the MySQL Workbench?

Key Terms • Review Questions • Project Questions

Appendix F: Getting Started with Microsoft Visio 2013

Chapter Objectives
What Is the Purpose of This Appendix?
Why Should I Learn to Use Microsoft Visio 2013?
What Will This Appendix Teach Me?
What Won't This Appendix Teach Me?
How Do I Start Microsoft Visio 2013?
How Do I Create a Database Model Diagram in Microsoft Visio 2013?
How Do I Name and Save a Database Model Diagram in Microsoft Visio 2013?
How Do I Create Entities in a Database Model Diagram in Microsoft Visio 2013?
How Do I Create Relationships Between Entities in a Data Model Diagram in Microsoft Visio 2013?
How Do I Create Diagrams Using Business Process Modeling Notation (BPMN) in Microsoft Visio 2013?

Key Terms • Review Questions • Project Questions

Appendix G: Data Structures for Database Processing

Chapter Objectives
Flat Files
Processing Flat Files in Multiple Orders • A Note on Record Addressing • Maintaining Order with Linked Lists • Maintaining Order with Indexes • B-Trees • Summary of Data Structures
Representing Binary Relationships
Review of Record Relationships • Representing Trees • Representing Simple Networks • Representing Complex Networks • Summary of Relationship Representations
Secondary-Key Representations
Linked-List Representation of Secondary Keys • Index Representation of Secondary Keys

Key Terms • Review Questions

Appendix H: The Semantic Object Model

Chapter Objectives
Semantic Objects
Defining Semantic Objects • *Attributes* • *Object Identifiers* • *Attribute Domains* • *Semantic Object Views*
Types of Objects
Simple Objects • *Composite Objects* • *Compound Objects* • *Representing One-to-One Compound Objects* • *Representing One-to-Many and Many-to-One Relationships* • *Representing Many-to-Many Relationships* • *Hybrid Objects* • *Representing Hybrid Relationships* • *Association Objects* • *Parent/Subtype Objects* • *Archetype/Version Objects*
Comparing the Semantic Object and the E-R Models

Key Terms • Review Questions

Appendix I: Getting Started with Web Servers, PHP, and the Eclipse PDT

Chapter Objectives
What Is the Purpose of This Appendix?
How Do I Install a Web Server?
How Do I Set Up IIS in Windows 7 and Windows 8?
How Do I Manage IIS in Windows 7 and Windows 8?
How Is a Web Site Structured?
How Do I View a Web Page from the IIS Web Server?
How Is Web Site Security Managed?
What Is the Eclipse PDT?
How Do I Install the Eclipse PDT?
What Is PHP?
How Do I Install PHP?
How Do I Create a Web Page Using the Eclipse PDT?
How Do I Manage the PHP Configuration?

Key Terms • Review Questions • Project Questions

Appendix J: Business Intelligence Systems

Chapter Objectives
What Is the Purpose of This Appendix?
Business Intelligence Systems
Reporting Systems and Data Mining Applications
Reporting Systems • *Data Mining Applications*
The Components of a Data Warehouse
Data Warehouses and Data Marts • *Data Warehouses and Dimensional Databases*
Reporting Systems
RFM Analysis • *Producing the RFM Report* • *Reporting System Components* • *Report Types* • *Report Media* • *Report Modes* • *Report System Functions* • *OLAP*
Data Mining
Unsupervised Data Mining • *Supervised Data Mining* • *Three Popular Data Mining Techniques* • *Market Basket Analysis*
Key Terms • Review Questions • Project Questions • Case Questions • The Queen Anne Curiosity Shop • Morgan Importing

Bibliography 613
Glossary 615
Index 629

Preface

The 13th edition of *Database Processing: Fundamentals, Design, and Implementation* refines the organization and content of this classic textbook to reflect a new teaching and professional workplace environment. Students and other readers of this book will benefit from new content and features in this edition.

New to This Edition

Content and features new to the 13th edition of *Database Processing: Fundamentals, Design, and Implementation* include:

- Material on *Big Data* and the evolving *NoSQL movement* has been moved to Chapter 12 and expanded upon. Big Data is the theme for the chapter. New material on virtualization, cloud computing, and the development of non-relational unstructured data stores (such as Cassandra and HBase) and the Hadoop Distributed File System (HDFS) is also included in Chapter 12.
- Each chapter now features an independent Case Question set. The Case Question sets are problem sets that generally do not require the student to have completed work on the same case in a previous chapter (there is one intentional exception that ties data modeling and database design together). Although in some instances the same basic named case may be used in different chapters, each instance is still completely independent of any other instance.
- The SQL topics of JOIN…ON and OUTER JOIN previously in Chapter 7 have been moved to Chapter 2 so nearly all SQL query topics are covered in one chapter (the exception is correlated subqueries, which are still reserved for Chapter 8).
- The coverage of SQL Persistent Stored Modules (SQL/PSM) in Chapter 7, Chapter 10, Chapter 10A, Chapter 10B, and Chapter 10C now includes a discussion of user-defined functions.
- The use of Microsoft Access 2013 to demonstrate and reinforce basic principles of database creation and use. This book has been revised to update all references to Microsoft Access and other Microsoft Office products (e.g., Microsoft Excel) to the recently released Microsoft Office 2013 versions.
- The updating of the book to reflect the use of Microsoft SQL Server 2012, the current version of Microsoft SQL Server. Although most of the topics covered are backward compatible with Microsoft SQL Server 2008 R2 and Microsoft SQL Server 2008 R2 Express edition, all material in the book now uses SQL Server 2012 in conjunction with Office 2013, exclusively.
- The updating of the book to use MySQL 5.6, which is the current generally available (GA) release of MySQL. Further, we also now use the MySQL Installer for Windows for installations on computers with the Windows operating system.
- The use of the Microsoft Windows Server 2012 as the server operating system and Windows 8 as the workstation operating system generally discussed and illustrated in the text. These are the current Microsoft server and workstation operating systems. We do keep some Windows 7 material where it seems appropriate and in some places present both the Windows 7 and Window 8 versions of operations and utilities.

- The addition of online Appendix J, "Business Intelligence Systems." This appendix contains material removed from Chapter 12 to make room for new material on Big Data and the Not Only SQL movement.
- We have updated online Appendix I, "Getting Started with Web Servers, PHP, and the Eclipse PDT." This new material provides a detailed introduction to the installation and use of the Microsoft IIS Web server, PHP, and the Eclipse IDE used for Web database application development as discussed in Chapter 11.

Fundamentals, Design, and Implementation

With today's technology, it is impossible to utilize a DBMS successfully without first learning fundamental concepts. After years of developing databases with business users, we have developed what we believe to be a set of essential database concepts. These are augmented by the concepts necessitated by the increasing use of the Internet, the World Wide Web, and commonly available analysis tools. Thus, the organization and topic selection of the 13th edition are designed to:

- Present an early introduction to SQL queries.
- Use a "spiral approach" to database design.
- Use a consistent, generic Information Engineering (IE) Crow's Foot E-R diagram notation for data modeling and database design.
- Provide a detailed discussion of specific normal forms within a discussion of normalization that focuses on pragmatic normalization techniques.
- Use current DBMS technology: Microsoft Access 2013, Microsoft SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6.
- Create Web database applications based on widely used Web development technology.
- Provide an introduction to business intelligence (BI) systems.
- Discuss the dimensional database concepts used in database designs for data warehouses and OnLine Analytical Processing (OLAP).
- Discuss the emerging and important topics of server virtualization, cloud computing, Big Data, and the Not Only SQL movement.

These changes have been made because it has become obvious that the basic structure of the earlier editions (up to and including the 9th edition—the 10th edition introduced many of the changes we used in the 11th and 12th editions and retain in the 13th edition) was designed for a teaching environment that no longer exists. The structural changes to the book were made for several reasons:

- Unlike the early years of database processing, today's students have ready access to data modeling and DBMS products.
- Today's students are too impatient to start a class with lengthy conceptual discussions on data modeling and database design. They want to do something, see a result, and obtain feedback.
- In the current economy, students need to reassure themselves that they are learning marketable skills.

Early Introduction of SQL DML

Given these changes in the classroom environment, this book provides an early introduction to SQL data manipulation language (DML) SELECT statements. The discussion of SQL data definition language (DDL) and additional DML statements occurs in Chapters 7 and 8. By encountering SQL SELECT statements in Chapter 2, students learn early in the class how to query data and obtain results, seeing firsthand some of the ways that database technology will be useful to them.

The text assumes that students will work through the SQL statements and examples with a DBMS product. This is practical today because nearly every student has access to Microsoft

Access. Therefore, Chapters 1 and 2 and Appendix A, "Getting Started with Microsoft Access 2013," are written to support an early introduction of Microsoft Access 2013 and the use of Microsoft Access 2013 for SQL queries (Microsoft Access 2013 QBE query techniques are also covered).

If a non–Microsoft Access-based approach is desired, versions of SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6 are readily available for use. Free versions of the three major DBMS products covered in this book (SQL Server 2012 Express, Oracle Database Express Edition 11*g* Release 2, and MySQL 5.6 Community Edition) are available for download. For a detailed discussion of the available DBMS products, see Chapter 10, pages 461-462. Thus, students can actively use a DBMS product by the end of the first week of class.

BY THE WAY

The presentation and discussion of SQL are spread over four chapters so students can learn about this important topic in small bites. SQL SELECT statements are taught in Chapter 2. SQL data definition language (DDL) and SQL data manipulation language (DML) statements are presented in Chapter 7. Correlated subqueries and EXISTS/NOT EXISTS statements are described in Chapter 8, while SQL transaction control language (TCL) and SQL data control language (DCL) are discussed in Chapter 9. Each topic appears in the context of accomplishing practical tasks. Correlated subqueries, for example, are used to verify functional dependency assumptions, a necessary task for database redesign.

This box illustrates another feature used in this book: BTW boxes are used to separate comments from the text discussion. Sometimes they present ancillary material; other times they reinforce important concepts.

A Spiral Approach to the Database Design Process

Today, databases arise from three sources: (1) from the need to integrate existing data from spreadsheets, data files, and database extracts; (2) from the need to develop new information systems to support business processes; and (3) from the need to redesign an existing database to adapt to changing requirements. We believe that the fact that these three sources exist present instructors with a significant pedagogical opportunity. Rather than teach database design just once from data models, why not teach database design three times, once for each of these sources? In practice, this idea has turned out to be even more successful than expected.

Database Design Iteration 1: Databases from Existing Data

Considering the design of databases from existing data, if someone were to e-mail us a set of tables and say, "Create a database from them," how would we proceed? We would examine the tables in light of normalization criteria and then determine whether the new database was for query only or whether it was for query and update. Depending on the answer, we would denormalize the data, joining them together, or we would normalize the data, pulling them apart. All of this is important for students to know and understand.

Therefore, the first iteration of database design gives instructors a rich opportunity to teach normalization, not as a set of theoretical concepts, but rather as a useful toolkit for making design decisions for databases created from existing data. Additionally, the construction of databases from existing data is an increasingly common task that is often assigned to junior staff members. Learning how to apply normalization to the design of databases from existing data not only provides an interesting way of teaching normalization, it is also common and useful!

We prefer to teach and use a pragmatic approach to normalization and present this approach in Chapter 3. However, we are aware that many instructors like to teach normalization in the context of a step-by-step normal form presentation (1NF, 2NF, 3NF, then BCNF), and Chapter 3 now includes additional material to provide more support for this approach as well.

In today's workplace, large organizations are increasingly licensing standardized software from vendors such as SAP, Oracle, and Siebel. Such software already has a database design. But with every organization running the same software, many are learning that they can only gain a competitive advantage if they make better use of the data in those predesigned databases. Hence, students who know how to extract data and create read-only databases for reporting

and data mining have obtained marketable skills in the world of ERP and other packaged software solutions.

Database Design Iteration 2: Data Modeling and Database Design

The second source of databases is from new systems development. Although not as common as in the past, many databases are still created from scratch. Thus, students still need to learn data modeling, and they still need to learn how to transform data models into database designs which are then implemented in a DBMS product.

The IE Crow's Foot Model as a Design Standard

This edition uses a generic, standard IE Crow's Foot notation. Your students should have no trouble understanding the symbols and using the data modeling or database design tool of your choice.

IDEF1X (which was used as the preferred E-R diagram notation in the 9th edition of this text) is explained in Appendix C, "E-R Diagrams and the IDEF1X Standard," in case your students will graduate into an environment where it is used or if you prefer to use it in your classes. UML is explained in Appendix D, "E-R Diagrams and the UML Standard," in case you prefer to use UML in your classes.

BY THE WAY

The choice of a data modeling tool is somewhat problematic. Of the two most readily available tools, Microsoft Visio 2013 has been rewritten as a very rudimentary database design tool, while Sun Microsystems MySQL Workbench, is a database design tool, not a data modeling tool. MySQL Workbench cannot produce an N:M relationship as such (as a data model requires), but has to immediately break it into two 1:N relationships (as database design does). Therefore, the intersection table must be constructed and modeled. This confounds data modeling with database design in just the way that we are attempting to teach students to avoid.

To be fair to Microsoft Visio 2013, it is true that data models with N:M relationships can be drawn using the standard Microsoft Visio 2013 drawing tools. Unfortunately, Microsoft has chosen to remove many of the best database design tools that were in Microsoft Visio 2010, and Microsoft Visio 2013 lacks the tools that made it a favorite of Microsoft Access and Microsoft SQL Server users. For a full discussion of these tools, see Appendix E, "Getting Started with the MySQL Workbench Data Modeling Tools," and Appendix F, "Getting Started with the Microsoft Visio 2013."

Good data modeling tools are available, but they tend to be more complex and expensive. Two examples are Visible Systems' Visible Analyst and CA Technologies' CA ERwin Data Modeler. Visible Analyst is available in a student edition (at a modest price), and a one-year time-limited CA ERwin Data Modeler Community Edition suitable for class use can be downloaded from *http://erwin.com/products/data-modeler/community-edition* has limited the number of objects that can be created by this edition to 25 entities per model and disabled some other features (see *http://erwin.com/content/products/CA-ERwin-r9-Community-Edition-Matrix-na.pdf*), but there is still enough functionality to make this product a possible choice for class use.

Database Design from E-R Data Models

As we discuss in Chapter 6, designing a database from data models consists of three tasks: representing entities and attributes with tables and columns; representing maximum cardinality by creating and placing foreign keys; and representing minimum cardinality via constraints, triggers, and application logic.

The first two tasks are straightforward. However, designs for minimum cardinality are more difficult. Required parents are easily enforced using NOT NULL foreign keys and referential integrity constraints. Required children are more problematic. In this book, however, we simplify the discussion of this topic by limiting the use of referential integrity actions and by supplementing those actions with design documentation. See the discussion around Figure 6-28.

Although the design for required children is complicated, it is important for students to learn. It also provides a reason for students to learn about triggers as well. In any case, the discussion of these topics is much simpler than it was in prior editions because of the use of the IE Crow's Foot model and ancillary design documentation.

Database Implementation from Database Designs

Of course, to complete the process, a database design must be implemented in a DBMS product. This is discussed in Chapter 7, where we introduce SQL DDL for creating tables and SQL DML for populating the tables with data.

David Kroenke is the creator of the semantic object model (SOM). The SOM is presented in Appendix H, "The Semantic Object Model." The E-R data model is used everywhere else in the text.

Database Design Iteration 3: Database Redesign

Database redesign, the third iteration of database design, is both common and difficult. As stated in Chapter 8, information systems cause organizational change. New information systems give users new behaviors, and as users behave in new ways, they require changes in their information systems.

Database redesign is by nature complex. Depending on your students, you may wish to skip it, and you can do so without loss of continuity. Database redesign is presented after the discussion of SQL DDL and DML in Chapter 7 because it requires the use of advanced SQL. It also provides a practical reason to teach correlated subqueries and EXISTS/NOT EXISTS statements.

Active Use of a DBMS Product

We assume that students will actively use a DBMS product. The only real question becomes "which one?" Realistically, most of us have four alternatives to consider: Microsoft Access, Microsoft SQL Server, Oracle Database, or MySQL. You can use any of those products with this text, and tutorials for each of them are presented for Microsoft Access 2013 (Appendix A), SQL Server 2012 (Chapter 10A), Oracle Database 11*g* Release 2 (Chapter 10B), and MySQL 5.6 (Chapter 10C). Given the limitations of class time, it is probably necessary to pick and use just one of these products. You can often devote a portion of a lecture to discussing the characteristics of each, but it is usually best to limit student work to one of them. The possible exception to this is starting the course with Microsoft Access and then switching to a more robust DBMS product later in the course.

Using Microsoft Access 2013

The primary advantage of Microsoft Access is accessibility. Most students already have a copy, and, if not, copies are easily obtained. Many students will have used Microsoft Access in their introductory or other classes. Appendix A, "Getting Started with Microsoft Access 2013," is a tutorial on Microsoft Access 2013 for students who have not used it but who wish to use it with this book.

However, Microsoft Access has several disadvantages. First, as explained in Chapter 1, Microsoft Access is a combination application generator and DBMS. Microsoft Access confuses students because it confounds database processing with application development. Also, Microsoft Access 2013 hides SQL behind its query processor and makes SQL appear as an afterthought rather than a foundation. Furthermore, as discussed in Chapter 2, Microsoft Access 2013 does not correctly process some of the basic SQL-92 standard statements in its default setup. Finally, Microsoft Access 2013 does not support triggers. You can simulate triggers by trapping Windows events, but that technique is nonstandard and does not effectively communicate the nature of trigger processing.

Using SQL Server 2012, Oracle Database 11*g* Release 2, or MySQL 5.6

Choosing which of these products to use depends on your local situation. Oracle Database 11*g* Release 2, a superb enterprise-class DBMS product, is difficult to install and administer.

However, if you have local staff to support your students, it can be an excellent choice. As shown in Chapter 10B, Oracle's SQL Developer GUI tool (or SQL*Plus if you are dedicated to this beloved command-line tool) is a handy tool for learning SQL, triggers, and stored procedures. In our experience, students require considerable support to install Oracle on their own computers, and you may be better off to use Oracle from a central server.

SQL Server 2012, although probably not as robust as Oracle Database 11*g* Release 2, is easy to install on Windows machines, and it provides the capabilities of an enterprise-class DBMS product. The standard database administrator tool is the Microsoft SQL Server Management Studio GUI tool. As shown in Chapter 10A, SQL Server 2012 can be used to learn SQL, triggers, and stored procedures.

MySQL 5.6, discussed in Chapter 10C, is an open-source DBMS product that is receiving increased attention and market share. The capabilities of MySQL are continually being upgraded, and MySQL 5.6 supports stored procedures and triggers. MySQL also has excellent GUI tools in the MySQL Workbench and an excellent command-line tool (the MySQL Command Line Client). It is the easiest of the three products for students to install on their own computers. It also works with the Linux operating system and is popular as part of the AMP (Apache–MySQL–PHP) package (known as WAMP on Windows and LAMP on Linux).

BY THE WAY

If the DBMS you use is not driven by local circumstances and you do have a choice, we recommend using SQL Server 2012. It has all of the features of an enterprise-class DBMS product, and it is easy to install and use. Another option is to start with Microsoft Access 2013 if it is available and switch to SQL Server 2012 at Chapter 7. Chapters 1 and 2 and Appendix A are written specifically to support this approach. A variant is to use Microsoft Access 2013 as the development tool for forms and reports running against an SQL Server 2012 database.

If you prefer a different DBMS product, you can still start with Microsoft Access 2013 and switch later in the course. See the detailed discussion of the available DBMS products on pages 461–462 in Chapter 10 for a good review of your options.

Focus on Database Application Processing

In this edition, we clearly draw the line between *application development* per se and *database application processing*. Specifically, we have:

- Focused on specific database dependent applications:
  - Web-based, database-driven applications
  - XML-based data processing
  - Business intelligence (BI) systems applications
- Emphasized the use of commonly available, multiple–OS-compatible application development languages.
- Limited the use of specialized vendor-specific tools and programming languages as much as possible.

There is simply not enough room in this book to provide even a basic introduction to programming languages used for application development such as the Microsoft .NET languages and Java. Therefore, rather than attempting to introduce these languages, we leave them for other classes where they can be covered at an appropriate depth. Instead, we focus on basic tools that are relatively straightforward to learn and immediately applicable to database-driven applications. We use PHP as our Web development language, and we use the readily available Eclipse integrate development environment (IDE) as our development tool. The result is a very focused final section of the book, where we deal specifically with the interface between databases and the applications that use them.

BY THE WAY

Although we try to use widely available software as much as possible, there are, of course, exceptions where we must use vendor-specific tools. For BI applications, for example, we draw on Microsoft Excel's PivotTable capabilities and the Microsoft PowerPivot for Microsoft Excel 2013 add-in, and on the Microsoft SQL Server 2012 SP1 Data Mining Add-ins for Microsoft Office. However, either alternatives to these tools are available (*OpenOffice.org* DataPilot capabilities, the Palo OLAP Server) or the tools are generally available for download.

Business Intelligence Systems and Dimensional Databases

This edition maintains coverage of business intelligence (BI) systems (Chapter 12 and Appendix J). The chapter includes a discussion of dimensional databases, which are the underlying structure for data warehouses, data marts, and OLAP servers. It still covers data management for data warehouses and data marts and also describes reporting and data mining applications, including OLAP.

Appendix J includes in-depth coverage of two applications that should be particularly interesting to students. The first is RFM analysis, a reporting application frequently used by mail order and e-commerce companies. The complete RFM analysis is accomplished in Appendix J through the use of standard SQL statements. Additionally, this chapter includes a market basket analysis that is also performed using SQL correlated subqueries. This chapter can be assigned at any point after Chapter 8 and could be used as a motivator to illustrate the practical applications of SQL midcourse.

Overview of the Chapters in the 13th Edition

Chapter 1 sets the stage by introducing database processing, describing basic components of database systems and summarizing the history of database processing. If students are using Microsoft Access 2013 for the first time (or need a good review), they will also need to study Appendix A, "Getting Started with Microsoft Access 2013," at this point. Chapter 2 presents SQL SELECT statements. It also includes sections on how to submit SQL statements to Microsoft Access 2013, SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6.

The next four chapters, Chapters 3 through 6, present the first two iterations of database design. Chapter 3 presents the principles of normalization to Boyce-Codd normal form (BNCF). It describes the problems of multivalued dependencies and explains how to eliminate them. This foundation in normalization is applied in Chapter 4 to the design of databases from existing data.

Chapters 5 and 6 describe the design of new databases. Chapter 5 presents the E-R data model. Traditional E-R symbols are explained, but the majority of the chapter uses IE Crow's Foot notation. Chapter 5 provides a taxonomy of entity types, including strong, ID-dependent, weak but not ID-dependent, supertype/subtype, and recursive. The chapter concludes with a simple modeling example for a university database.

Chapter 6 describes the transformation of data models into database designs by converting entities and attributes to tables and columns, by representing maximum cardinality by creating and placing foreign keys, and by representing minimum cardinality via carefully designed DBMS constraints, triggers, and application code. The primary section of this chapter parallels the entity taxonomy in Chapter 5.

Chapter 7 presents SQL DDL, DML, and SQL/Persistent Stored Modules (SQL/PSM). SQL DDL is used to implement the design of an example introduced in Chapter 6. INSERT, UPDATE, MERGE, and DELETE statements are discussed, as are SQL views. Additionally, the principles of embedding SQL in program code are presented, SQL/PSM is discussed, and triggers and stored procedures are explained.

Database redesign, the third iteration of database design, is described in Chapter 8. This chapter presents SQL correlated subqueries and EXISTS/NOT EXISTS statements and uses those statements in the redesign process. Reverse engineering is described, and basic redesign patterns are illustrated and discussed.

Chapters 9, 10, 10A, 10B, and 10C consider the management of multiuser organizational databases. Chapter 9 describes database administration tasks, including concurrency, security, and backup and recovery. Chapter 10 is a general introduction to the online Chapters 10A, 10B, and 10C, which describe SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6, respectively. These chapters show how to use these products to create database structures and process SQL statements. They also explain concurrency, security, and backup and recovery with each product. The discussion in Chapters 10A, 10B, and 10C parallels the order of discussion in Chapter 9 as much as possible, though rearrangements of some topics are made, as needed, to support the discussion of a specific DBMS product.

BY THE WAY

We have maintained or extended our coverage of Microsoft Access, Microsoft SQL Server, Oracle Database, and MySQL (introduced in *Database Processing: Fundamentals, Design, and Implementation,* 11th edition) in this book. In order to keep the bound book to a reasonable length and to keep the cost of the book down, we have chosen to provide some material by download from our Web site at *www.pearsoninternationaleditions.com/kroenke*. There you will find:

- Chapter 10A—Managing Databases with SQL Server 2012
- Chapter 10B—Managing Databases with Oracle Database 11*g* Release 2
- Chapter 10C—Managing Databases with MySQL 5.6
- Appendix A—Getting Started with Microsoft Access 2013
- Appendix B—Getting Started with Systems Analysis and Design
- Appendix C—E-R Diagrams and the IDEF1X Standard
- Appendix D—E-R Diagrams and the UML Standard
- Appendix E—Getting Started with MySQL Workbench Data Modeling Tools
- Appendix F—Getting Started with Microsoft Visio 2013
- Appendix G—Data Structures for Database Processing
- Appendix H—The Semantic Object Model
- Appendix I—Getting Started with Web Servers, PHP, and the Eclipse PDT
- Appendix J—Business Intelligence Systems

Chapters 11 and 12 address standards for accessing databases. Chapter 11 presents ODBC, OLE DB, ADO.NET, ASP.NET, JDBC, and JavaServer Pages (JSP). It then introduces PHP (and the Eclipse IDE) and illustrates the use of PHP for the publication of databases via Web pages. This is followed by a description of the integration of XML and database technology. The chapter begins with a primer on XML and then shows how to use the FOR XML SQL statement in SQL Server.

Chapter 12 concludes the text with a discussion of BI systems, dimensional data models, data warehouses, data marts, server virtualization, cloud computing, Big Data, structured storage, and the Not Only SQL movement.

Supplements

This text is accompanied by a wide variety of supplements. Please visit the text's Web site at *www.pearsoninternationaleditions.com/kroenke* to access the instructor and student supplements described below. Please contact your Pearson sales representative for more details. All supplements were written by David Auer and Robert Crossler.

For Students

- Many of the sample databases used in this text are available online in Microsoft Access, SQL Server 2012, Oracle Database 11*g* Release 2, and MySQL 5.6 format.

For Instructors

- The *Instructor's Resource Manual* provides sample course syllabi, teaching suggestions, and answers to end-of-chapter review, project, and case questions.
- The *Test Item File* and *TestGen* include an extensive set of test questions in multiple-choice, true/false, fill-in-the-blank, short-answer, and essay format. The difficulty level and where the topic is covered in the text are noted for each question. The Test Item File is available in Microsoft Word and in TestGen. The TestGen software is PC/MAC compatible and preloaded with all of the Test Item File questions. You can manually or randomly view test questions and drag and drop to create a test. You can add or modify test-bank questions as needed. Our TestGens are converted for use in BlackBoard, WebCT, Angel, D2L, and Moodle. All conversions are available on the IRC.
- *PowerPoint Presentation Slides* feature lecture notes that highlight key terms and concepts. Instructors can customize the presentation by adding their own slides or editing the existing ones.
- The *Image Library* is a collection of the text art organized by chapter. This includes all figures, tables, and screenshots (as permission allows) to enhance class lectures and PowerPoint presentations.

Acknowledgments

We are grateful for the support of many people in the development of this 13th edition and previous editions. Thanks to Rick Mathieu at James Madison University for interesting and insightful discussions on the database course. Professor Doug MacLachlan from the Marketing Department at the University of Washington was most helpful in understanding the goals, objectives, and technology of data mining, particularly as it pertains to marketing. Don Nilson, formerly of the Microsoft Corporation, helped us understand the importance of XML to database processing. Kraig Pencil and Jon Junell of the College of Business and Economics at Western Washington University helped us refine the use of the book in the classroom.

In addition, we wish to thank the reviewers of this and previous editions:

Ann Aksut, *Central Piedmont Community College*
Allen Badgett, *Oklahoma City University*
Rich Beck, *Washington University*
Jeffrey J. Blessing, *Milwaukee School of Engineering*
Alan Brandyberry, *Kent State University*
Larry Booth, *Clayton State University*
Jason Deane, *Virginia Polytechnic Institute and State University*
Barry Flaschbart, *Missouri University of Science and Technology*
Andy Green, *Kennesaw State University*
Dianne Hall, *Auburn University*
Jeff Hassett, *University of Utah*
Barbara Hewitt, *Texas A&M, Kingsville*
William Hochstettler, *Franklin University*
Margaret Hvatum, *St. Louis Community College*
Nitin Kale, *University of Southern California, Los Angeles*
Darrel Karbginsky, *Chemeketa Community College*
Johnny Li, *South University*
Lin Lin, *New Jersey Institute of Technology*
Mike Morris, *Southeastern Oklahoma State University*
Jane Perschbach, *Texas A&M University–Central Texas*
Catherine Ricardo, *Iona College*
Kevin Roberts, *DeVry University*

Ioulia Rytikova, *George Mason University*
Christelle Scharff, *Pace University*
Julian M. Scher, *New Jersey Institute of Technology*
Namchul Shin, *Pace University*
K. David Smith, *Cameron University*
M. Jane Stafford, *Columbia College–Jefferson City*
Marcia Williams, *Bellevue Community College*
Timothy Woodcock, *Texas A&M University–Central Texas*

Finally, we would like to thank Bob Horan, our editor; Kelly Loftus, our editorial project manager during this project; Jane Bonnell, our production project manager; and Angel Chavez, our project manager; for their professionalism, insight, support, and assistance in the development of this project. We would also like to thank Robert Crossler for his detailed comments on the final manuscript. Finally, David Kroenke would like to thank his wife, Lynda, and David Auer would like to thank his wife, Donna, for their love, encouragement, and patience while this project was being completed.

David Kroenke
Seattle, Washington

David Auer
Bellingham, Washington

The publishers wish to thank Arup Kumar Bhattacharjee and Soumen Mukherjee for reviewing the content of the International Edition.

About the Authors

David M. Kroenke

Work Experience

David M. Kroenke has more than 35 years' experience in the computer industry. He began as a computer programmer for the U.S. Air Force, working both in Los Angeles and at the Pentagon, where he developed one of the world's first DBMS products while part of a team that created a computer simulation of World War III. That simulation served a key role for strategic weapons studies during a 10-year period of the Cold War.

From 1973 to 1978, Kroenke taught in the College of Business at Colorado State University. In 1977, he published the first edition of *Database Processing*, a significant and successful textbook that, more than 30 years later, you now are reading in its 13th edition. In 1978, he left Colorado State and joined Boeing Computer Services, where he managed the team that designed database management components of the IPAD project. After that, he joined with Steve Mitchell to form Mitchell Publishing and worked as an editor and author, developing texts, videos, and other educational products and seminars. Mitchell Publishing was acquired by Random House in 1986. During those years, he also worked as an independent consultant, primarily as a database disaster repairman helping companies recover from failed database projects.

In 1982, Kroenke was one of the founding directors of the Microrim Corporation. From 1984 to 1987, he served as the Vice President of Product Marketing and Development and managed the team that created and marketed the DBMS product R:base 5000 as well as other related products.

For the next five years, Kroenke worked independently while he developed a new data modeling language called the *semantic object model*. He licensed this technology to the Wall Data Corporation in 1992 and then served as the Chief Technologist for Wall Data's SALSA line of products. He was awarded three software patents on this technology.

Since 1998, Kroenke has continued consulting and writing. His current interests concern the practical applications of data mining techniques on large organizational databases. An avid sailor, he wrote *Know Your Boat: The Guide to Everything That Makes Your Boat Work*, which was published by McGraw-Hill in 2002.

Consulting

Kroenke has consulted with numerous organizations during his career. In 1978, he worked for Fred Brooks, consulting with IBM on a project that became the DBMS product DB2. In 1989, he consulted for the Microsoft Corporation on a project that became Microsoft Access. In the 1990s, he worked with Computer Sciences Corporation and with General Research Corporation for the development of technology and products that were used to model all of the U.S. Army's logistical data as part of the CALS project. Additionally, he has consulted for Boeing Computer Services, the U.S. Air Force Academy, Logicon Corporation, and other smaller organizations.

Publications

- *Database Processing*, Pearson Prentice Hall, 13 editions, 1977–present (coauthor with David Auer, 11th, 12th, and 13th editions)
- *Database Concepts*, Pearson Prentice Hall, six editions, 2004–present (coauthor with David Auer, 3rd, 4th, 5th, and 6th editions)