The background of the cover is a close-up photograph of a wood grain, showing various shades of brown and tan with prominent vertical and diagonal lines. The text is overlaid on this background.

FUNDAMENTALS OF  
DATABASE  
SYSTEMS

7<sup>TH</sup> Edition

ELMASRI • NAVATHE

FUNDAMENTALS OF

# Database Systems

SEVENTH EDITION

This page intentionally left blank

FUNDAMENTALS OF

# Database Systems

SEVENTH EDITION

**Ramez Elmasri**

*Department of Computer Science and Engineering  
The University of Texas at Arlington*

**Shamkant B. Navathe**

*College of Computing  
Georgia Institute of Technology*

**PEARSON**

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



Vice President and Editorial Director, ECS:  
*Marcia J. Horton*  
Acquisitions Editor: *Matt Goldstein*  
Editorial Assistant: *Kelsey Loanes*  
Marketing Managers: *Bram Van Kempen, Demetrius Hall*  
Marketing Assistant: *Jon Bryant*  
Senior Managing Editor: *Scott Disanno*  
Production Project Manager: *Rose Kernan*  
Program Manager: *Carole Snyder*  
Global HE Director of Vendor Sourcing  
and Procurement: *Diane Hynes*  
Director of Operations: *Nick Sklitsis*

Operations Specialist: *Maura Zaldivar-Garcia*  
Cover Designer: *Black Horse Designs*  
Manager, Rights and Permissions: *Rachel Youdelman*  
Associate Project Manager, Rights and Permissions:  
*Timothy Nicholls*  
Full-Service Project Management: *Rashmi Tickyani,*  
*iEnergizer Aptara®, Ltd.*  
Composition: *iEnergizer Aptara®, Ltd.*  
Printer/Binder: *Edwards Brothers Malloy*  
Cover Printer: *Phoenix Color/Hagerstown*  
Cover Image: *Micha Pawlitzki/Terra/Corbis*  
Typeface: *10.5/12 Minion Pro*

**Copyright © 2016, 2011, 2007 by Ramez Elmasri and Shamkant B. Navathe.** All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright and permissions should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use materials from this work, please submit a written request to Pearson Higher Education, Permissions Department, 221 River Street, Hoboken, NJ 07030.

Many of the designations by manufacturers and seller to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages with, or arising out of, the furnishing, performance, or use of these programs.

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.

#### **Library of Congress Cataloging-in-Publication Data on File**

10 9 8 7 6 5 4 3 2 1

**PEARSON**

ISBN-10: 0-13-397077-9

ISBN-13: 978-0-13-397077-7

*To Amalia  
and  
to Ramy, Riyad, Katrina, and Thomas  
R. E.*

*To my wife Aruna for her love, support, and understanding  
and  
to Rohan, Maya, and Ayush for bringing so much joy into our lives  
S.B.N.*

This page intentionally left blank

# Preface

This book introduces the fundamental concepts necessary for designing, using, and implementing database systems and database applications. Our presentation stresses the fundamentals of database modeling and design, the languages and models provided by the database management systems, and database system implementation techniques. The book is meant to be used as a textbook for a one- or two-semester course in database systems at the junior, senior, or graduate level, and as a reference book. Our goal is to provide an in-depth and up-to-date presentation of the most important aspects of database systems and applications, and related technologies. We assume that readers are familiar with elementary programming and data-structuring concepts and that they have had some exposure to the basics of computer organization.

## New to This Edition

The following key features have been added in the seventh edition:

- A reorganization of the chapter ordering (this was based on a survey of the instructors who use the textbook); however, the book is still organized so that the individual instructor can choose to follow the new chapter ordering or *choose a different ordering of chapters* (for example, follow the chapter order from the sixth edition) when presenting the materials.
- There are two new chapters on recent advances in database systems and big data processing; one new chapter (Chapter 24) covers an introduction to the newer class of database systems known as **NOSQL databases**, and the other new chapter (Chapter 25) covers technologies for processing **big data**, including **MapReduce** and **Hadoop**.
- The chapter on query processing and optimization has been expanded and reorganized into two chapters; Chapter 18 focuses on strategies and algorithms for query processing whereas Chapter 19 focuses on query optimization techniques.
- A second UNIVERSITY database example has been added to the early chapters (Chapters 3 through 8) in addition to our COMPANY database example from the previous editions.
- Many of the individual chapters have been updated to varying degrees to include newer techniques and methods; rather than discuss these enhancements here,



we will describe them later in the preface when we discuss the organization of the seventh edition.

The following are key features of the book:

- A self-contained, flexible organization that can be tailored to individual needs; in particular, *the chapters can be used in different orders* depending on the instructor's preference.
- A companion website (<http://www.pearsonhighered.com/cs-resources>) includes data to be loaded into various types of relational databases for more realistic student laboratory exercises.
- A dependency chart (shown later in this preface) to show which chapters depend on other earlier chapters; this can guide the instructor who wants to tailor the *order of presentation of the chapters*.
- A collection of supplements, including a robust set of materials for instructors and students such as PowerPoint slides, figures from the text, and an instructor's guide with solutions.

## Organization and Contents of the Seventh Edition

There are some organizational changes in the seventh edition as well as improvement to the individual chapters. The book is now divided into 12 parts as follows:

- Part 1 (Chapters 1 and 2) describes the basic introductory concepts necessary for a good understanding of database models, systems, and languages. Chapters 1 and 2 introduce databases, typical users, and DBMS concepts, terminology, and architecture, as well as a discussion of the progression of database technologies over time and a brief history of data models. These chapters have been updated to introduce some of the newer technologies such as NOSQL systems.
- Part 2 (Chapters 3 and 4) includes the presentation on entity-relationship modeling and database design; however, it is *important to note* that instructors can cover the relational model chapters (Chapters 5 through 8) *before Chapters 3 and 4* if that is their preferred order of presenting the course materials. In Chapter 3, the concepts of the Entity-Relationship (ER) model and ER diagrams are presented and used to illustrate conceptual database design. Chapter 4 shows how the basic ER model can be extended to incorporate additional modeling concepts such as subclasses, specialization, generalization, union types (categories) and inheritance, leading to the enhanced-ER (EER) data model and EER diagrams. The notation for the class diagrams of UML are also introduced in Chapters 7 and 8 as an alternative model and diagrammatic notation for ER/EER diagrams.
- Part 3 (Chapters 5 through 8) includes a detailed presentation on relational databases and SQL with some additional new material in the SQL chapters to cover a few SQL constructs that were not in the previous edition. Chapter 5

describes the basic relational model, its integrity constraints, and update operations. Chapter 6 describes some of the basic parts of the SQL standard for relational databases, including data definition, data modification operations, and simple SQL queries. Chapter 7 presents more complex SQL queries, as well as the SQL concepts of triggers, assertions, views, and schema modification. Chapter 8 describes the formal operations of the relational algebra and introduces the relational calculus. The material on SQL (Chapters 6 and 7) is presented before our presentation on relational algebra and calculus in Chapter 8 to allow instructors to start SQL projects early in a course if they wish (it is possible to cover Chapter 8 before Chapters 6 and 7 if the instructor desires this order). The final chapter in Part 2, Chapter 9, covers ER- and EER-to-relational mapping, which are algorithms that can be used for designing a relational database schema from a conceptual ER/EER schema design.

- Part 4 (Chapters 10 and 11) are the chapters on database programming techniques; these chapters can be assigned as reading materials and augmented with materials on the particular language used in the course for programming projects (much of this documentation is readily available on the Web). Chapter 10 covers traditional SQL programming topics, such as embedded SQL, dynamic SQL, ODBC, SQLJ, JDBC, and SQL/CLI. Chapter 11 introduces Web database programming, using the PHP scripting language in our examples, and includes new material that discusses Java technologies for Web database programming.
- Part 5 (Chapters 12 and 13) covers the updated material on object-relational and object-oriented databases (Chapter 12) and XML (Chapter 13); both of these chapters now include a presentation of how the SQL standard incorporates object concepts and XML concepts into more recent versions of the SQL standard. Chapter 12 first introduces the concepts for object databases, and then shows how they have been incorporated into the SQL standard in order to add object capabilities to relational database systems. It then covers the ODMG object model standard, and its object definition and query languages. Chapter 13 covers the XML (eXtensible Markup Language) model and languages, and discusses how XML is related to database systems. It presents XML concepts and languages, and compares the XML model to traditional database models. We also show how data can be converted between the XML and relational representations, and the SQL commands for extracting XML documents from relational tables.
- Part 6 (Chapters 14 and 15) are the normalization and relational design theory chapters (we moved all the formal aspects of normalization algorithms to Chapter 15). Chapter 14 defines functional dependencies, and the normal forms that are based on functional dependencies. Chapter 14 also develops a step-by-step intuitive normalization approach, and includes the definitions of multivalued dependencies and join dependencies. Chapter 15 covers normalization theory, and the formalisms, theories,

and algorithms developed for relational database design by normalization, including the relational decomposition algorithms and the relational synthesis algorithms.

- Part 7 (Chapters 16 and 17) contains the chapters on file organizations on disk (Chapter 16) and indexing of database files (Chapter 17). Chapter 16 describes primary methods of organizing files of records on disk, including ordered (sorted), unordered (heap), and hashed files; both static and dynamic hashing techniques for disk files are covered. Chapter 16 has been updated to include materials on buffer management strategies for DBMSs as well as an overview of new storage devices and standards for files and modern storage architectures. Chapter 17 describes indexing techniques for files, including B-tree and B<sup>+</sup>-tree data structures and grid files, and has been updated with new examples and an enhanced discussion on indexing, including how to choose appropriate indexes and index creation during physical design.
- Part 8 (Chapters 18 and 19) includes the chapters on query processing algorithms (Chapter 18) and optimization techniques (Chapter 19); these two chapters have been updated and reorganized from the single chapter that covered both topics in the previous editions and include some of the newer techniques that are used in commercial DBMSs. Chapter 18 presents algorithms for searching for records on disk files, and for joining records from two files (tables), as well as for other relational operations. Chapter 18 contains new material, including a discussion of the semi-join and anti-join operations with examples of how they are used in query processing, as well as a discussion of techniques for selectivity estimation. Chapter 19 covers techniques for query optimization using cost estimation and heuristic rules; it includes new material on nested subquery optimization, use of histograms, physical optimization, and join ordering methods and optimization of typical queries in data warehouses.
- Part 9 (Chapters 20, 21, and 22) covers transaction processing concepts; concurrency control; and database recovery from failures. These chapters have been updated to include some of the newer techniques that are used in some commercial and open source DBMSs. Chapter 20 introduces the techniques needed for transaction processing systems, and defines the concepts of recoverability and serializability of schedules; it has a new section on buffer replacement policies for DBMSs and a new discussion on the concept of snapshot isolation. Chapter 21 gives an overview of the various types of concurrency control protocols, with a focus on two-phase locking. We also discuss timestamp ordering and optimistic concurrency control techniques, as well as multiple-granularity locking. Chapter 21 includes a new presentation of concurrency control methods that are based on the snapshot isolation concept. Finally, Chapter 23 focuses on database recovery protocols, and gives an overview of the concepts and techniques that are used in recovery.

- Part 10 (Chapters 23, 24, and 25) includes the chapter on distributed databases (Chapter 23), plus the two new chapters on NOSQL storage systems for big data (Chapter 24) and big data technologies based on Hadoop and MapReduce (Chapter 25). Chapter 23 introduces distributed database concepts, including availability and scalability, replication and fragmentation of data, maintaining data consistency among replicas, and many other concepts and techniques. In Chapter 24, NOSQL systems are categorized into four general categories with an example system in each category used for our examples, and the data models, operations, as well as the replication/distribution/scalability strategies of each type of NOSQL system are discussed and compared. In Chapter 25, the MapReduce programming model for distributed processing of big data is introduced, and then we have presentations of the Hadoop system and HDFS (Hadoop Distributed File System), as well as the Pig and Hive high-level interfaces, and the YARN architecture.
- Part 11 (Chapters 26 through 29) is entitled Advanced Database Models, Systems, and Applications and includes the following materials: Chapter 26 introduces several advanced data models including active databases/triggers (Section 26.1), temporal databases (Section 26.2), spatial databases (Section 26.3), multimedia databases (Section 26.4), and deductive databases (Section 26.5). Chapter 27 discusses information retrieval (IR) and Web search, and includes topics such as IR and keyword-based search, comparing DB with IR, retrieval models, search evaluation, and ranking algorithms. Chapter 28 is an introduction to data mining including overviews of various data mining methods such as associate rule mining, clustering, classification, and sequential pattern discovery. Chapter 29 is an overview of data warehousing including topics such as data warehousing models and operations, and the process of building a data warehouse.
- Part 12 (Chapter 30) includes one chapter on database security, which includes a discussion of SQL commands for discretionary access control (GRANT, REVOKE), as well as mandatory security levels and models for including mandatory access control in relational databases, and a discussion of threats such as SQL injection attacks, as well as other techniques and methods related to data security and privacy.

Appendix A gives a number of alternative diagrammatic notations for displaying a conceptual ER or EER schema. These may be substituted for the notation we use, if the instructor prefers. Appendix B gives some important physical parameters of disks. Appendix C gives an overview of the QBE graphical query language, and Appendixes D and E (available on the book's Companion Website located at <http://www.pearsonhighered.com/elmasri>) cover legacy database systems, based on the hierarchical and network database models. They have been used for more than thirty years as a basis for many commercial database applications and transaction-processing systems.

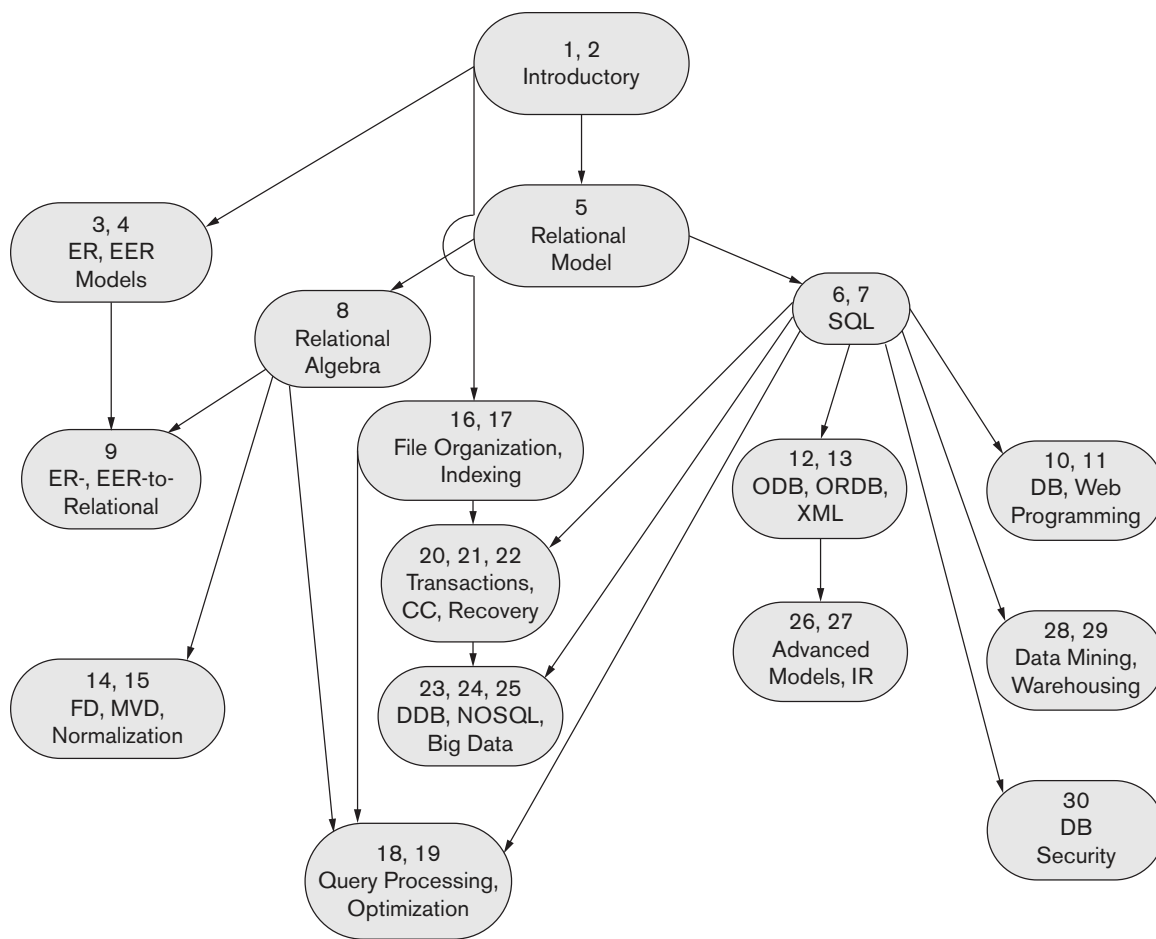
## Guidelines for Using This Book

There are many different ways to teach a database course. The chapters in Parts 1 through 7 can be used in an introductory course on database systems in the order that they are given or in the preferred order of individual instructors. Selected chapters and sections may be left out and the instructor can add other chapters from the rest of the book, depending on the emphasis of the course. At the end of the opening section of some of the book's chapters, we list sections that are candidates for being left out whenever a less-detailed discussion of the topic is desired. We suggest covering up to Chapter 15 in an introductory database course and including selected parts of other chapters, depending on the background of the students and the desired coverage. For an emphasis on system implementation techniques, chapters from Parts 7, 8, and 9 should replace some of the earlier chapters.

Chapters 3 and 4, which cover conceptual modeling using the ER and EER models, are important for a good conceptual understanding of databases. However, they may be partially covered, covered later in a course, or even left out if the emphasis is on DBMS implementation. Chapters 16 and 17 on file organizations and indexing may also be covered early, later, or even left out if the emphasis is on database models and languages. For students who have completed a course on file organization, parts of these chapters can be assigned as reading material or some exercises can be assigned as a review for these concepts.

If the emphasis of a course is on database design, then the instructor should cover Chapters 3 and 4 early on, followed by the presentation of relational databases. A total life-cycle database design and implementation project would cover conceptual design (Chapters 3 and 4), relational databases (Chapters 5, 6, and 7), data model mapping (Chapter 9), normalization (Chapter 14), and application programs implementation with SQL (Chapter 10). Chapter 11 also should be covered if the emphasis is on Web database programming and applications. Additional documentation on the specific programming languages and RDBMS used would be required. The book is written so that it is possible to cover topics in various sequences. The following chapter dependency chart shows the major dependencies among chapters. As the diagram illustrates, it is possible to start with several different topics following the first two introductory chapters. Although the chart may seem complex, it is important to note that if the chapters are covered in order, the dependencies are not lost. The chart can be consulted by instructors wishing to use an alternative order of presentation.

For a one-semester course based on this book, selected chapters can be assigned as reading material. The book also can be used for a two-semester course sequence. The first course, *Introduction to Database Design and Database Systems*, at the sophomore, junior, or senior level, can cover most of Chapters 1 through 15. The second course, *Database Models and Implementation Techniques*, at the senior or first-year graduate level, can cover most of Chapters 16 through 30. The two-semester sequence can also be designed in various other ways, depending on the preferences of the instructors.



## Supplemental Materials

Support material is available to qualified instructors at Pearson's instructor resource center (<http://www.pearsonhighered.com/irc>). For access, contact your local Pearson representative.

- PowerPoint lecture notes and figures.
- A solutions manual.

## Acknowledgments

It is a great pleasure to acknowledge the assistance and contributions of many individuals to this effort. First, we would like to thank our editor, Matt Goldstein, for his guidance, encouragement, and support. We would like to acknowledge the excellent work of Rose Kernan for production management, Patricia Daly for a



thorough copy editing of the book, Martha McMaster for her diligence in proofing the pages, and Scott Disanno, Managing Editor of the production team. We also wish to thank Kelsey Loanes from Pearson for her continued help with the project, and reviewers Michael Doherty, Deborah Dunn, Imad Rahal, Karen Davis, Gilliean Lee, Leo Mark, Monisha Pulimood, Hassan Reza, Susan Vrbsky, Li Da Xu, Weining Zhang and Vincent Oria.

Ramez Elmasri would like to thank Kulsawasd Jitkajornwanich, Vivek Sharma, and Surya Swaminathan for their help with preparing some of the material in Chapter 24. Sham Navathe would like to acknowledge the following individuals who helped in critically reviewing and revising various topics. Dan Forsythe and Satish Damle for discussion of storage systems; Rafi Ahmed for detailed re-organization of the material on query processing and optimization; Harish Butani, Balaji Palanisamy, and Prajakta Kalmegh for their help with the Hadoop and MapReduce technology material; Vic Ghorpadey and Nenad Jukic for revision of the Data Warehousing material; and finally, Frank Rietta for newer techniques in database security, Kunal Malhotra for various discussions, and Saurav Sahay for advances in information retrieval systems.

We would like to repeat our thanks to those who have reviewed and contributed to previous editions of *Fundamentals of Database Systems*.

- **First edition.** Alan Apt (editor), Don Batory, Scott Downing, Dennis Heimbinger, Julia Hodges, Yannis Ioannidis, Jim Larson, Per-Ake Larson, Dennis McLeod, Rahul Patel, Nicholas Roussopoulos, David Stemple, Michael Stonebraker, Frank Tompa, and Kyu-Young Whang.
- **Second edition.** Dan Joraanstad (editor), Rafi Ahmed, Antonio Albano, David Beech, Jose Blakeley, Panos Chrysanthis, Suzanne Dietrich, Vic Ghorpadey, Goetz Graefe, Eric Hanson, Junguk L. Kim, Roger King, Vram Kouramajian, Vijay Kumar, John Lowther, Sanjay Manchanda, Toshimi Minoura, Inderpal Mumick, Ed Omiecinski, Girish Pathak, Raghu Ramakrishnan, Ed Robertson, Eugene Sheng, David Stotts, Marianne Winslett, and Stan Zdonick.
- **Third edition.** Maite Suarez-Rivas and Katherine Harutunian (editors); Suzanne Dietrich, Ed Omiecinski, Rafi Ahmed, Francois Bancilhon, Jose Blakeley, Rick Cattell, Ann Chervenak, David W. Embley, Henry A. Etlinger, Leonidas Fegaras, Dan Forsyth, Farshad Fotouhi, Michael Franklin, Sreejith Gopinath, Goetz Craefe, Richard Hull, Sushil Jajodia, Ramesh K. Karne, Harish Kotbagi, Vijay Kumar, Tarcisio Lima, Ramon A. Mata-Toledo, Jack McCaw, Dennis McLeod, Rokia Missaoui, Magdi Morsi, M. Narayanaswamy, Carlos Ordonez, Joan Peckham, Betty Salzberg, Ming-Chien Shan, Junping Sun, Rajshekhar Sunderraman, Aravindan Veerasamy, and Emilia E. Villareal.
- **Fourth edition.** Maite Suarez-Rivas, Katherine Harutunian, Daniel Rausch, and Juliet Silveri (editors); Phil Bernhard, Zhengxin Chen, Jan Chomicki, Hakan Ferhatosmanoglu, Len Fisk, William Hankley, Ali R. Hurson, Vijay Kumar, Peretz Shoval, Jason T. L. Wang (reviewers); Ed Omiecinski (who contributed to Chapter 27). Contributors from the University of Texas at

Arlington are Jack Fu, Hyoil Han, Babak Hojabri, Charley Li, Ande Swathi, and Steven Wu; Contributors from Georgia Tech are Weimin Feng, Dan Forsythe, Angshuman Guin, Abrar Ul-Haque, Bin Liu, Ying Liu, Wanxia Xie, and Waigen Yee.

- **Fifth edition.** Matt Goldstein and Katherine Harutunian (editors); Michelle Brown, Gillian Hall, Patty Mahtani, Maite Suarez-Rivas, Bethany Tidd, and Joyce Cosentino Wells (from Addison-Wesley); Hani Abu-Salem, Jamal R. Alsabbagh, Ramzi Bualuan, Soon Chung, Sumali Conlon, Hasan Davulcu, James Geller, Le Gruenwald, Latifur Khan, Herman Lam, Byung S. Lee, Donald Sanderson, Jamil Saquer, Costas Tsatsoulis, and Jack C. Wileden (reviewers); Raj Sunderraman (who contributed the laboratory projects); Salman Azar (who contributed some new exercises); Gaurav Bhatia, Fariborz Farahmand, Ying Liu, Ed Omiecinski, Nalini Polavarapu, Liora Sahar, Saurav Sahay, and Wanxia Xie (from Georgia Tech).
- **Sixth edition.** Matt Goldstein (editor); Gillian Hall (production management); Rebecca Greenberg (copy editing); Jeff Holcomb, Marilyn Lloyd, Margaret Waples, and Chelsea Bell (from Pearson); Rafi Ahmed, Venu Dasigi, Neha Deodhar, Fariborz Farahmand, Hariprasad Kumar, Leo Mark, Ed Omiecinski, Balaji Palanisamy, Nalini Polavarapu, Parimala R. Pranesh, Bharath Rengarajan, Liora Sahar, Saurav Sahay, Narsi Srinivasan, and Wanxia Xie.

Last, but not least, we gratefully acknowledge the support, encouragement, and patience of our families.

*R. E.*

*S.B.N.*

This page intentionally left blank

# Contents

Preface vii

About the Authors xxx

## ■ part 1

### Introduction to Databases ■

#### chapter 1 Databases and Database Users 3

- 1.1 Introduction 4
- 1.2 An Example 6
- 1.3 Characteristics of the Database Approach 10
- 1.4 Actors on the Scene 15
- 1.5 Workers behind the Scene 17
- 1.6 Advantages of Using the DBMS Approach 17
- 1.7 A Brief History of Database Applications 23
- 1.8 When Not to Use a DBMS 27
- 1.9 Summary 27
- Review Questions 28
- Exercises 28
- Selected Bibliography 29

#### chapter 2 Database System Concepts and Architecture 31

- 2.1 Data Models, Schemas, and Instances 32
- 2.2 Three-Schema Architecture and Data Independence 36
- 2.3 Database Languages and Interfaces 38
- 2.4 The Database System Environment 42
- 2.5 Centralized and Client/Server Architectures for DBMSs 46
- 2.6 Classification of Database Management Systems 51
- 2.7 Summary 54
- Review Questions 55
- Exercises 55
- Selected Bibliography 56

## ■ part 2

### Conceptual Data Modeling and Database Design ■

#### chapter 3 Data Modeling Using the Entity–Relationship (ER) Model 59

- 3.1 Using High-Level Conceptual Data Models for Database Design 60
- 3.2 A Sample Database Application 62
- 3.3 Entity Types, Entity Sets, Attributes, and Keys 63
- 3.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints 72
- 3.5 Weak Entity Types 79
- 3.6 Refining the ER Design for the COMPANY Database 80
- 3.7 ER Diagrams, Naming Conventions, and Design Issues 81
- 3.8 Example of Other Notation: UML Class Diagrams 85
- 3.9 Relationship Types of Degree Higher than Two 88
- 3.10 Another Example: A UNIVERSITY Database 92
- 3.11 Summary 94
- Review Questions 96
- Exercises 96
- Laboratory Exercises 103
- Selected Bibliography 104

#### chapter 4 The Enhanced Entity–Relationship (EER) Model 107

- 4.1 Subclasses, Superclasses, and Inheritance 108
- 4.2 Specialization and Generalization 110
- 4.3 Constraints and Characteristics of Specialization and Generalization Hierarchies 113
- 4.4 Modeling of UNION Types Using Categories 120
- 4.5 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions 122
- 4.6 Example of Other Notation: Representing Specialization and Generalization in UML Class Diagrams 127
- 4.7 Data Abstraction, Knowledge Representation, and Ontology Concepts 128
- 4.8 Summary 135
- Review Questions 135
- Exercises 136
- Laboratory Exercises 143
- Selected Bibliography 146

## ■ part **3**

### The Relational Data Model and SQL ■

chapter <b>5</b>	<b>The Relational Data Model and Relational Database Constraints</b>	<b>149</b>
5.1	Relational Model Concepts	150
5.2	Relational Model Constraints and Relational Database Schemas	157
5.3	Update Operations, Transactions, and Dealing with Constraint Violations	165
5.4	Summary	169
	Review Questions	170
	Exercises	170
	Selected Bibliography	175
chapter <b>6</b>	<b>Basic SQL</b>	<b>177</b>
6.1	SQL Data Definition and Data Types	179
6.2	Specifying Constraints in SQL	184
6.3	Basic Retrieval Queries in SQL	187
6.4	INSERT, DELETE, and UPDATE Statements in SQL	198
6.5	Additional Features of SQL	201
6.6	Summary	202
	Review Questions	203
	Exercises	203
	Selected Bibliography	205
chapter <b>7</b>	<b>More SQL: Complex Queries, Triggers, Views, and Schema Modification</b>	<b>207</b>
7.1	More Complex SQL Retrieval Queries	207
7.2	Specifying Constraints as Assertions and Actions as Triggers	225
7.3	Views (Virtual Tables) in SQL	228
7.4	Schema Change Statements in SQL	232
7.5	Summary	234
	Review Questions	236
	Exercises	236
	Selected Bibliography	238
chapter <b>8</b>	<b>The Relational Algebra and Relational Calculus</b>	<b>239</b>
8.1	Unary Relational Operations: SELECT and PROJECT	241
8.2	Relational Algebra Operations from Set Theory	246



8.3 Binary Relational Operations: JOIN and DIVISION	251
8.4 Additional Relational Operations	259
8.5 Examples of Queries in Relational Algebra	265
8.6 The Tuple Relational Calculus	268
8.7 The Domain Relational Calculus	277
8.8 Summary	279
Review Questions	280
Exercises	281
Laboratory Exercises	286
Selected Bibliography	288

## chapter **9** Relational Database Design by ER- and EER-to-Relational Mapping 289

9.1 Relational Database Design Using ER-to-Relational Mapping	290
9.2 Mapping EER Model Constructs to Relations	298
9.3 Summary	303
Review Questions	303
Exercises	303
Laboratory Exercises	305
Selected Bibliography	306

# ■ part 4

## Database Programming Techniques ■

### chapter **10** Introduction to SQL Programming Techniques 309

10.1 Overview of Database Programming Techniques and Issues	310
10.2 Embedded SQL, Dynamic SQL, and SQLJ	314
10.3 Database Programming with Function Calls and Class Libraries: SQL/CLI and JDBC	326
10.4 Database Stored Procedures and SQL/PSM	335
10.5 Comparing the Three Approaches	338
10.6 Summary	339
Review Questions	340
Exercises	340
Selected Bibliography	341

### chapter **11** Web Database Programming Using PHP 343

11.1 A Simple PHP Example	344
11.2 Overview of Basic Features of PHP	346

11.3 Overview of PHP Database Programming	353
11.4 Brief Overview of Java Technologies for Database Web Programming	358
11.5 Summary	358
Review Questions	359
Exercises	359
Selected Bibliography	359

## ■ part 5

### Object, Object-Relational, and XML: Concepts, Models, Languages, and Standards ■

#### chapter **12** Object and Object-Relational Databases 363

12.1 Overview of Object Database Concepts	365
12.2 Object Database Extensions to SQL	379
12.3 The ODMG Object Model and the Object Definition Language ODL	386
12.4 Object Database Conceptual Design	405
12.5 The Object Query Language OQL	408
12.6 Overview of the C++ Language Binding in the ODMG Standard	417
12.7 Summary	418
Review Questions	420
Exercises	421
Selected Bibliography	422

#### chapter **13** XML: Extensible Markup Language 425

13.1 Structured, Semistructured, and Unstructured Data	426
13.2 XML Hierarchical (Tree) Data Model	430
13.3 XML Documents, DTD, and XML Schema	433
13.4 Storing and Extracting XML Documents from Databases	442
13.5 XML Languages	443
13.6 Extracting XML Documents from Relational Databases	447
13.7 XML/SQL: SQL Functions for Creating XML Data	453
13.8 Summary	455
Review Questions	456
Exercises	456
Selected Bibliography	456

## ■ part 6

### Database Design Theory and Normalization ■

#### chapter **14** Basics of Functional Dependencies and Normalization for Relational Databases 459

- 14.1 Informal Design Guidelines for Relation Schemas 461
- 14.2 Functional Dependencies 471
- 14.3 Normal Forms Based on Primary Keys 474
- 14.4 General Definitions of Second and Third Normal Forms 483
- 14.5 Boyce-Codd Normal Form 487
- 14.6 Multivalued Dependency and Fourth Normal Form 491
- 14.7 Join Dependencies and Fifth Normal Form 494
- 14.8 Summary 495
- Review Questions 496
- Exercises 497
- Laboratory Exercises 501
- Selected Bibliography 502

#### chapter **15** Relational Database Design Algorithms and Further Dependencies 503

- 15.1 Further Topics in Functional Dependencies: Inference Rules, Equivalence, and Minimal Cover 505
- 15.2 Properties of Relational Decompositions 513
- 15.3 Algorithms for Relational Database Schema Design 519
- 15.4 About Nulls, Dangling Tuples, and Alternative Relational Designs 523
- 15.5 Further Discussion of Multivalued Dependencies and 4NF 527
- 15.6 Other Dependencies and Normal Forms 530
- 15.7 Summary 533
- Review Questions 534
- Exercises 535
- Laboratory Exercises 536
- Selected Bibliography 537

## ■ part 7

### File Structures, Hashing, Indexing, and Physical Database Design ■

#### chapter **16** Disk Storage, Basic File Structures, Hashing, and Modern Storage Architectures 541

- 16.1 Introduction 542
- 16.2 Secondary Storage Devices 547
- 16.3 Buffering of Blocks 556
- 16.4 Placing File Records on Disk 560
- 16.5 Operations on Files 564
- 16.6 Files of Unordered Records (Heap Files) 567
- 16.7 Files of Ordered Records (Sorted Files) 568
- 16.8 Hashing Techniques 572
- 16.9 Other Primary File Organizations 582
- 16.10 Parallelizing Disk Access Using RAID Technology 584
- 16.11 Modern Storage Architectures 588
- 16.12 Summary 592
- Review Questions 593
- Exercises 595
- Selected Bibliography 598

#### chapter **17** Indexing Structures for Files and Physical Database Design 601

- 17.1 Types of Single-Level Ordered Indexes 602
- 17.2 Multilevel Indexes 613
- 17.3 Dynamic Multilevel Indexes Using B-Trees and B<sup>+</sup>-Trees 617
- 17.4 Indexes on Multiple Keys 631
- 17.5 Other Types of Indexes 633
- 17.6 Some General Issues Concerning Indexing 638
- 17.7 Physical Database Design in Relational Databases 643
- 17.8 Summary 646
- Review Questions 647
- Exercises 648
- Selected Bibliography 650

## ■ part 8

### Query Processing and Optimization ■

#### chapter **18** Strategies for Query Processing 655

- 18.1 Translating SQL Queries into Relational Algebra and Other Operators 657
- 18.2 Algorithms for External Sorting 660
- 18.3 Algorithms for SELECT Operation 663
- 18.4 Implementing the JOIN Operation 668
- 18.5 Algorithms for PROJECT and Set Operations 676
- 18.6 Implementing Aggregate Operations and Different Types of JOINS 678
- 18.7 Combining Operations Using Pipelining 681
- 18.8 Parallel Algorithms for Query Processing 683
- 18.9 Summary 688
- Review Questions 688
- Exercises 689
- Selected Bibliography 689

#### chapter **19** Query Optimization 691

- 19.1 Query Trees and Heuristics for Query Optimization 692
- 19.2 Choice of Query Execution Plans 701
- 19.3 Use of Selectivities in Cost-Based Optimization 710
- 19.4 Cost Functions for SELECT Operation 714
- 19.5 Cost Functions for the JOIN Operation 717
- 19.6 Example to Illustrate Cost-Based Query Optimization 726
- 19.7 Additional Issues Related to Query Optimization 728
- 19.8 An Example of Query Optimization in Data Warehouses 731
- 19.9 Overview of Query Optimization in Oracle 733
- 19.10 Semantic Query Optimization 737
- 19.11 Summary 738
- Review Questions 739
- Exercises 740
- Selected Bibliography 740

## ■ part 9

### Transaction Processing, Concurrency Control, and Recovery ■

#### chapter **20** Introduction to Transaction Processing Concepts and Theory 745

- 20.1 Introduction to Transaction Processing 746
- 20.2 Transaction and System Concepts 753
- 20.3 Desirable Properties of Transactions 757
- 20.4 Characterizing Schedules Based on Recoverability 759
- 20.5 Characterizing Schedules Based on Serializability 763
- 20.6 Transaction Support in SQL 773
- 20.7 Summary 776
- Review Questions 777
- Exercises 777
- Selected Bibliography 779

#### chapter **21** Concurrency Control Techniques 781

- 21.1 Two-Phase Locking Techniques for Concurrency Control 782
- 21.2 Concurrency Control Based on Timestamp Ordering 792
- 21.3 Multiversion Concurrency Control Techniques 795
- 21.4 Validation (Optimistic) Techniques and Snapshot Isolation Concurrency Control 798
- 21.5 Granularity of Data Items and Multiple Granularity Locking 800
- 21.6 Using Locks for Concurrency Control in Indexes 805
- 21.7 Other Concurrency Control Issues 806
- 21.8 Summary 807
- Review Questions 808
- Exercises 809
- Selected Bibliography 810

#### chapter **22** Database Recovery Techniques 813

- 22.1 Recovery Concepts 814
- 22.2 NO-UNDO/REDO Recovery Based on Deferred Update 821
- 22.3 Recovery Techniques Based on Immediate Update 823



22.4 Shadow Paging	826
22.5 The ARIES Recovery Algorithm	827
22.6 Recovery in Multidatabase Systems	831
22.7 Database Backup and Recovery from Catastrophic Failures	832
22.8 Summary	833
Review Questions	834
Exercises	835
Selected Bibliography	838

## ■ part 10

### Distributed Databases, NOSQL Systems, and Big Data ■

#### chapter **23** Distributed Database Concepts 841

23.1 Distributed Database Concepts	842
23.2 Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design	847
23.3 Overview of Concurrency Control and Recovery in Distributed Databases	854
23.4 Overview of Transaction Management in Distributed Databases	857
23.5 Query Processing and Optimization in Distributed Databases	859
23.6 Types of Distributed Database Systems	865
23.7 Distributed Database Architectures	868
23.8 Distributed Catalog Management	875
23.9 Summary	876
Review Questions	877
Exercises	878
Selected Bibliography	880

#### chapter **24** NOSQL Databases and Big Data Storage Systems 883

24.1 Introduction to NOSQL Systems	884
24.2 The CAP Theorem	888
24.3 Document-Based NOSQL Systems and MongoDB	890
24.4 NOSQL Key-Value Stores	895
24.5 Column-Based or Wide Column NOSQL Systems	900
24.6 NOSQL Graph Databases and Neo4j	903
24.7 Summary	909
Review Questions	909
Selected Bibliography	910

## chapter **25** Big Data Technologies Based on MapReduce and Hadoop 911

- 25.1 What Is Big Data? 914
- 25.2 Introduction to MapReduce and Hadoop 916
- 25.3 Hadoop Distributed File System (HDFS) 921
- 25.4 MapReduce: Additional Details 926
- 25.5 Hadoop v2 alias YARN 936
- 25.6 General Discussion 944
- 25.7 Summary 953
- Review Questions 954
- Selected Bibliography 956

## ■ part **11**

### Advanced Database Models, Systems, and Applications ■

## chapter **26** Enhanced Data Models: Introduction to Active, Temporal, Spatial, Multimedia, and Deductive Databases 961

- 26.1 Active Database Concepts and Triggers 963
- 26.2 Temporal Database Concepts 974
- 26.3 Spatial Database Concepts 987
- 26.4 Multimedia Database Concepts 994
- 26.5 Introduction to Deductive Databases 999
- 26.6 Summary 1012
- Review Questions 1014
- Exercises 1015
- Selected Bibliography 1018

## chapter **27** Introduction to Information Retrieval and Web Search 1021

- 27.1 Information Retrieval (IR) Concepts 1022
- 27.2 Retrieval Models 1029
- 27.3 Types of Queries in IR Systems 1035
- 27.4 Text Preprocessing 1037
- 27.5 Inverted Indexing 1040
- 27.6 Evaluation Measures of Search Relevance 1044
- 27.7 Web Search and Analysis 1047

27.8 Trends in Information Retrieval	1057
27.9 Summary	1063
Review Questions	1064
Selected Bibliography	1066

chapter **28** **Data Mining Concepts** 1069

28.1 Overview of Data Mining Technology	1070
28.2 Association Rules	1073
28.3 Classification	1085
28.4 Clustering	1088
28.5 Approaches to Other Data Mining Problems	1091
28.6 Applications of Data Mining	1094
28.7 Commercial Data Mining Tools	1094
28.8 Summary	1097
Review Questions	1097
Exercises	1098
Selected Bibliography	1099

chapter **29** **Overview of Data Warehousing and OLAP** 1101

29.1 Introduction, Definitions, and Terminology	1102
29.2 Characteristics of Data Warehouses	1103
29.3 Data Modeling for Data Warehouses	1105
29.4 Building a Data Warehouse	1111
29.5 Typical Functionality of a Data Warehouse	1114
29.6 Data Warehouse versus Views	1115
29.7 Difficulties of Implementing Data Warehouses	1116
29.8 Summary	1117
Review Questions	1117
Selected Bibliography	1118

■ **part 12**

**Additional Database Topics: Security** ■

chapter **30** **Database Security** 1121

30.1 Introduction to Database Security Issues	1122
30.2 Discretionary Access Control Based on Granting and Revoking Privileges	1129
30.3 Mandatory Access Control and Role-Based Access Control for Multilevel Security	1134

30.4 SQL Injection	1143
30.5 Introduction to Statistical Database Security	1146
30.6 Introduction to Flow Control	1147
30.7 Encryption and Public Key Infrastructures	1149
30.8 Privacy Issues and Preservation	1153
30.9 Challenges to Maintaining Database Security	1154
30.10 Oracle Label-Based Security	1155
30.11 Summary	1158
Review Questions	1159
Exercises	1160
Selected Bibliography	1161

appendix **A** **Alternative Diagrammatic Notations for ER Models** 1163

appendix **B** **Parameters of Disks** 1167

appendix **C** **Overview of the QBE Language** 1171

C.1 Basic Retrievals in QBE 1171

C.2 Grouping, Aggregation, and Database Modification in QBE 1175

appendix **D** **Overview of the Hierarchical Data Model**

(located on the Companion Website at  
<http://www.pearsonhighered.com/elmasri>)

appendix **E** **Overview of the Network Data Model**

(located on the Companion Website at  
<http://www.pearsonhighered.com/elmasri>)

**Selected Bibliography** 1179

**Index** 1215