Sixth Edition

Fundamentals of Database Systems

Elmasri • Navathe

FUNDAMENTALS OF

Database Systems

SIXTH EDITION

This page intentionally left blank

FUNDAMENTALS OF

Database Systems

SIXTH 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

Addison-Wesley

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

Editor in Chief: *Michael Hirsch*
Acquisitions Editor: *Matt Goldstein*
Editorial Assistant: *Chelsea Bell*
Managing Editor: *Jeffrey Holcomb*
Senior Production Project Manager: *Marilyn Lloyd*
Media Producer: *Katelyn Boller*
Director of Marketing: *Margaret Waples*
Marketing Coordinator: *Kathryn Ferranti*
Senior Manufacturing Buyer: *Alan Fischer*
Senior Media Buyer: *Ginny Michaud*
Text Designer: *Sandra Rigney and Gillian Hall*
Cover Designer: *Elena Sidorova*
Cover Image: *Lou Gibbs/Getty Images*
Full Service Vendor: *Gillian Hall, The Aardvark Group*
Copyeditor: *Rebecca Greenberg*
Proofreader: *Holly McLean-Aldis*
Indexer: *Jack Lewis*
Printer/Binder: *Courier, Westford*
Cover Printer: *Lehigh-Phoenix Color/Hagerstown*

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

The interior of this book was set in Minion and Akzidenz Grotesk.

Copyright © 2011, 2007, 2004, 2000, 1994, and 1989 Pearson Education, Inc., publishing as Addison-Wesley. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright, and permission 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 material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, 501 Boylston Street, Suite 900, Boston, Massachusetts 02116.

Many of the designations by manufacturers and sellers 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.

Library of Congress Cataloging-in-Publication Data

Elmasri, Ramez.
Fundamentals of database systems / Ramez Elmasri, Shamkant B. Navathe.—6th ed.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-136-08620-8
1. Database management. I. Navathe, Sham. II. Title.

QA76.9.D3E57 2010
005.74—dc22

Addison-Wesley
is an imprint of

PEARSON

10 9 8 7 6 5 4 3 2 1—CW—14 13 12 11 10
ISBN 10: 0-136-08620-9
ISBN 13: 978-0-136-08620-8

To Katrina, Thomas, and Dora
(and also to Ficky)

R. E.

To my wife Aruna, mother Vijaya,
and to my entire family
for their love and support

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 sixth edition:

- A reorganization of the chapter ordering to allow instructors to start with projects and laboratory exercises very early in the course
- The material on SQL, the relational database standard, has been moved early in the book to Chapters 4 and 5 to allow instructors to focus on this important topic at the beginning of a course
- The material on object-relational and object-oriented databases has been updated to conform to the latest SQL and ODMG standards, and consolidated into a single chapter (Chapter 11)
- The presentation of XML has been expanded and updated, and moved earlier in the book to Chapter 12
- The chapters on normalization theory have been reorganized so that the first chapter (Chapter 15) focuses on intuitive normalization concepts, while the second chapter (Chapter 16) focuses on the formal theories and normalization algorithms
- The presentation of database security threats has been updated with a discussion on SQL injection attacks and prevention techniques in Chapter 24, and an overview of label-based security with examples

- Our presentation on spatial databases and multimedia databases has been expanded and updated in Chapter 26
- A new Chapter 27 on information retrieval techniques has been added, which discusses models and techniques for retrieval, querying, browsing, and indexing of information from Web documents; we present the typical processing steps in an information retrieval system, the evaluation metrics, and how information retrieval techniques are related to databases and to Web search

The following are key features of the book:

- A self-contained, flexible organization that can be tailored to individual needs
- A Companion Website (http://www.aw.com/elmasri) includes data to be loaded into various types of relational databases for more realistic student laboratory exercises
- A simple relational algebra and calculus interpreter
- 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 of the Sixth Edition

There are significant organizational changes in the sixth edition, as well as improvement to the individual chapters. The book is now divided into eleven parts as follows:

- Part 1 (Chapters 1 and 2) includes the introductory chapters
- The presentation on relational databases and SQL has been moved to Part 2 (Chapters 3 through 6) of the book; Chapter 3 presents the formal relational model and relational database constraints; the material on SQL (Chapters 4 and 5) is now presented before our presentation on relational algebra and calculus in Chapter 6 to allow instructors to start SQL projects early in a course if they wish (this reordering is also based on a study that suggests students master SQL better when it is taught before the formal relational languages)
- The presentation on entity-relationship modeling and database design is now in Part 3 (Chapters 7 through 10), but it can still be covered before Part 2 if the focus of a course is on database design
- Part 4 covers the updated material on object-relational and object-oriented databases (Chapter 11) and XML (Chapter 12)
- Part 5 includes the chapters on database programming techniques (Chapter 13) and Web database programming using PHP (Chapter 14, which was moved earlier in the book)
- Part 6 (Chapters 15 and 16) are the normalization and design theory chapters (we moved all the formal aspects of normalization algorithms to Chapter 16)

- Part 7 (Chapters 17 and 18) contains the chapters on file organizations, indexing, and hashing
- Part 8 includes the chapters on query processing and optimization techniques (Chapter 19) and database tuning (Chapter 20)
- Part 9 includes Chapter 21 on transaction processing concepts; Chapter 22 on concurrency control; and Chapter 23 on database recovery from failures
- Part 10 on additional database topics includes Chapter 24 on database security and Chapter 25 on distributed databases
- Part 11 on advanced database models and applications includes Chapter 26 on advanced data models (active, temporal, spatial, multimedia, and deductive databases); the new Chapter 27 on information retrieval and Web search; and the chapters on data mining (Chapter 28) and data warehousing (Chapter 29)

Contents of the Sixth Edition

Part 1 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.

Part 2 describes the relational data model, the SQL standard, and the formal relational languages. Chapter 3 describes the basic relational model, its integrity constraints, and update operations. Chapter 4 describes some of the basic parts of the SQL standard for relational databases, including data definition, data modification operations, and simple SQL queries. Chapter 5 presents more complex SQL queries, as well as the SQL concepts of triggers, assertions, views, and schema modification. Chapter 6 describes the operations of the relational algebra and introduces the relational calculus.

Part 3 covers several topics related to conceptual database modeling and database design. In Chapter 7, the concepts of the Entity-Relationship (ER) model and ER diagrams are presented and used to illustrate conceptual database design. Chapter 8 focuses on data abstraction and semantic data modeling concepts and shows how the ER model can be extended to incorporate these ideas, leading to the enhanced-ER (EER) data model and EER diagrams. The concepts presented in Chapter 8 include subclasses, specialization, generalization, and union types (categories). The notation for the class diagrams of UML is also introduced in Chapters 7 and 8. Chapter 9 discusses relational database design using ER- and EER-to-relational mapping. We end Part 3 with Chapter 10, which presents an overview of the different phases of the database design process in enterprises for medium-sized and large database applications.

Part 4 covers the object-oriented, object-relational, and XML data models, and their affiliated languages and standards. Chapter 11 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 12 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.

Part 5 is on database programming techniques. Chapter 13 covers SQL programming topics, such as embedded SQL, dynamic SQL, ODBC, SQLJ, JDBC, and SQL/CLI. Chapter 14 introduces Web database programming, using the PHP scripting language in our examples.

Part 6 covers normalization theory. Chapters 15 and 16 cover the formalisms, theories, and algorithms developed for relational database design by normalization. This material includes functional and other types of dependencies and normal forms of relations. Step-by-step intuitive normalization is presented in Chapter 15, which also defines multivalued and join dependencies. Relational design algorithms based on normalization, along with the theoretical materials that the algorithms are based on, are presented in Chapter 16.

Part 7 describes the physical file structures and access methods used in database systems. Chapter 17 describes primary methods of organizing files of records on disk, including static and dynamic hashing. Chapter 18 describes indexing techniques for files, including B-tree and B^+-tree data structures and grid files.

Part 8 focuses on query processing and database performance tuning. Chapter 19 introduces the basics of query processing and optimization, and Chapter 20 discusses physical database design and tuning.

Part 9 discusses transaction processing, concurrency control, and recovery techniques, including discussions of how these concepts are realized in SQL. Chapter 21 introduces the techniques needed for transaction processing systems, and defines the concepts of recoverability and serializability of schedules. Chapter 22 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. Finally, Chapter 23 focuses on database recovery protocols, and gives an overview of the concepts and techniques that are used in recovery.

Parts 10 and 11 cover a number of advanced topics. Chapter 24 gives an overview of database security including the discretionary access control model with SQL commands to GRANT and REVOKE privileges, the mandatory access control model with user categories and polyinstantiation, a discussion of data privacy and its relationship to security, and an overview of SQL injection attacks. Chapter 25 gives an introduction to distributed databases and discusses the three-tier client/server architecture. Chapter 26 introduces several enhanced database models for advanced applications. These include active databases and triggers, as well as temporal, spatial, multimedia, and deductive databases. Chapter 27 is a new chapter on information retrieval techniques, and how they are related to database systems and to Web

search methods. Chapter 28 on data mining gives an overview of the process of data mining and knowledge discovery, discusses algorithms for association rule mining, classification, and clustering, and briefly covers other approaches and commercial tools. Chapter 29 introduces data warehousing and OLAP concepts.

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. Appendixes D and E (available on the book's Companion Website located at http://www.aw.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. We consider it important to expose database management students to these legacy approaches so they can gain a better insight of how database technology has progressed.

Guidelines for Using This Book

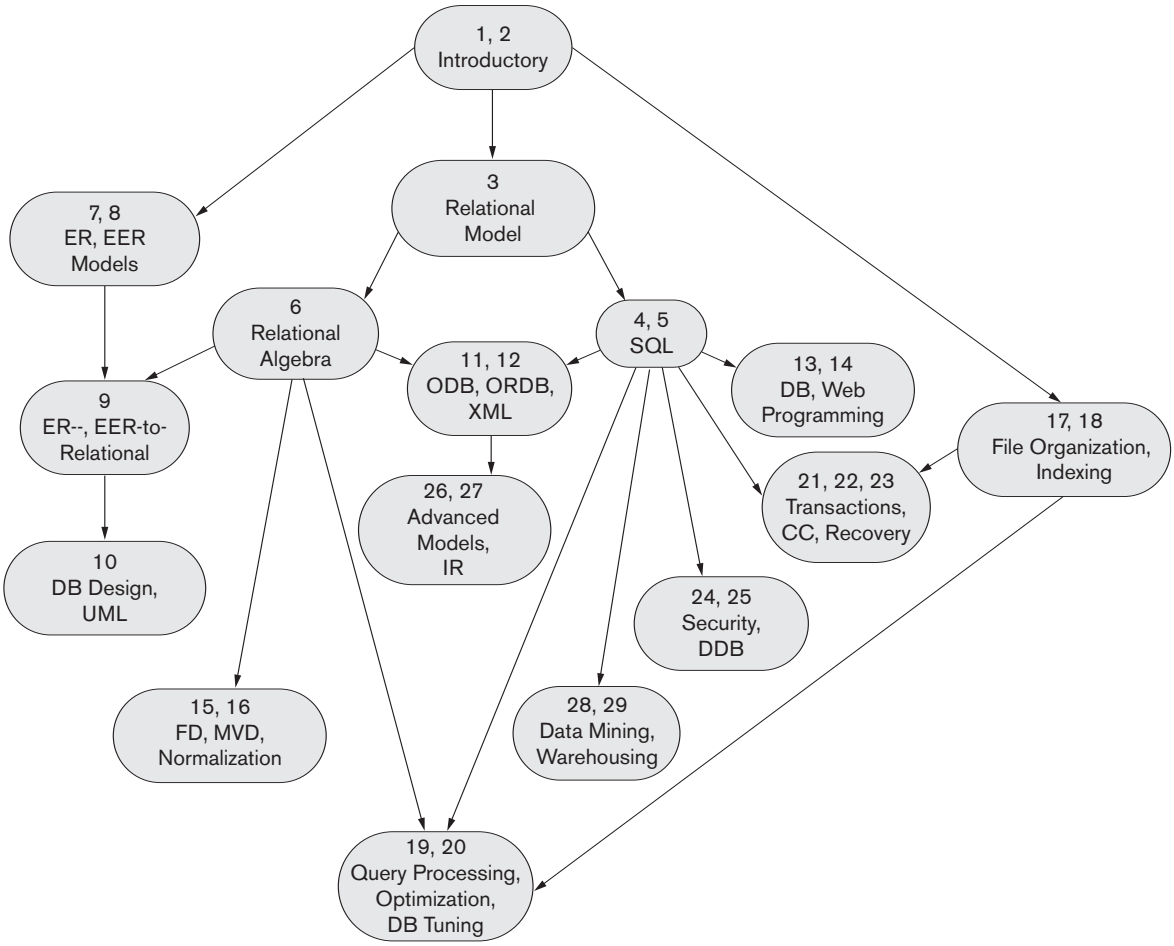
There are many different ways to teach a database course. The chapters in Parts 1 through 7 can be used in an introductory course on database systems in the order that they are given or in the preferred order of individual instructors. Selected chapters and sections may be left out, and the instructor 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 many 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 7 and 8, 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 17 and 18 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 7 and 8 early on, followed by the presentation of relational databases. A total life-cycle database design and implementation project would cover conceptual design (Chapters 7 and 8), relational databases (Chapters 3, 4, and 5), data model mapping (Chapter 9), normalization (Chapter 15), and application programs implementation with SQL (Chapter 13). Chapter 14 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 chapter dependency chart below 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 29. The two-semester sequence can also been designed in various other ways, depending on the preferences of the instructors.



Supplemental Materials

Support material is available to all users of this book and additional material is available to qualified instructors.

- PowerPoint lecture notes and figures are available at the Computer Science support Website at http://www.aw.com/cssupport.
- A lab manual for the sixth edition is available through the Companion Website (http://www.aw.com/elmasri). The lab manual contains coverage of popular data modeling tools, a relational algebra and calculus interpreter, and examples from the book implemented using two widely available database management systems. Select end-of-chapter laboratory problems in the book are correlated to the lab manual.
- A solutions manual is available to qualified instructors. Visit Addison-Wesley's instructor resource center (http://www.aw.com/irc), contact your local Addison-Wesley sales representative, or e-mail computing@aw.com for information about how to access the solutions.

Additional Support Material

Gradiance, an online homework and tutorial system that provides additional practice and tests comprehension of important concepts, is available to U.S. adopters of this book. For more information, please e-mail computing@aw.com or contact your local Pearson representative.

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 Gillian Hall for production management and Rebecca Greenberg for a thorough copy editing of the book. We thank the following persons from Pearson who have contributed to the sixth edition: Jeff Holcomb, Marilyn Lloyd, Margaret Waples, and Chelsea Bell.

Sham Navathe would like to acknowledge the significant contribution of Saurav Sahay to Chapter 27. Several current and former students also contributed to various chapters in this edition: Rafi Ahmed, Liora Sahar, Fariborz Farahmand, Nalini Polavarapu, and Wanxia Xie (former students); and Bharath Rengarajan, Narsi Srinivasan, Parimala R. Pranesh, Neha Deodhar, Balaji Palanisamy and Hariprasad Kumar (current students). Discussions with his colleagues Ed Omiecinski and Leo Mark at Georgia Tech and Venu Dasigi at SPSU, Atlanta have also contributed to the revision of the material.

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).

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

R. E.

S.B.N.

Contents

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 9
1.4 Actors on the Scene 14
1.5 Workers behind the Scene 16
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 26
1.9 Summary 27
Review Questions 27
Exercises 28
Selected Bibliography 28

chapter 2 **Database System Concepts and Architecture** 29

2.1 Data Models, Schemas, and Instances 30
2.2 Three-Schema Architecture and Data Independence 33
2.3 Database Languages and Interfaces 36
2.4 The Database System Environment 40
2.5 Centralized and Client/Server Architectures for DBMSs 44
2.6 Classification of Database Management Systems 49
2.7 Summary 52
Review Questions 53
Exercises 54
Selected Bibliography 55

part 2
The Relational Data Model and SQL

chapter 3 The Relational Data Model and Relational Database Constraints 59

3.1 Relational Model Concepts 60
3.2 Relational Model Constraints and Relational Database Schemas 67
3.3 Update Operations, Transactions, and Dealing with Constraint Violations 75
3.4 Summary 79
Review Questions 80
Exercises 80
Selected Bibliography 85

chapter 4 Basic SQL 87

4.1 SQL Data Definition and Data Types 89
4.2 Specifying Constraints in SQL 94
4.3 Basic Retrieval Queries in SQL 97
4.4 INSERT, DELETE, and UPDATE Statements in SQL 107
4.5 Additional Features of SQL 110
4.6 Summary 111
Review Questions 112
Exercises 112
Selected Bibliography 114

chapter 5 More SQL: Complex Queries, Triggers, Views, and Schema Modification 115

5.1 More Complex SQL Retrieval Queries 115
5.2 Specifying Constraints as Assertions and Actions as Triggers 131
5.3 Views (Virtual Tables) in SQL 133
5.4 Schema Change Statements in SQL 137
5.5 Summary 139
Review Questions 141
Exercises 141
Selected Bibliography 143

chapter 6 The Relational Algebra and Relational Calculus 145

6.1 Unary Relational Operations: SELECT and PROJECT 147
6.2 Relational Algebra Operations from Set Theory 152
6.3 Binary Relational Operations: JOIN and DIVISION 157
6.4 Additional Relational Operations 165
6.5 Examples of Queries in Relational Algebra 171
6.6 The Tuple Relational Calculus 174
6.7 The Domain Relational Calculus 183
6.8 Summary 185
Review Questions 186
Exercises 187
Laboratory Exercises 192
Selected Bibliography 194

part 3
Conceptual Modeling and Database Design

chapter 7 Data Modeling Using the Entity-Relationship (ER) Model 199

7.1 Using High-Level Conceptual Data Models for Database Design 200
7.2 A Sample Database Application 202
7.3 Entity Types, Entity Sets, Attributes, and Keys 203
7.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints 212
7.5 Weak Entity Types 219
7.6 Refining the ER Design for the COMPANY Database 220
7.7 ER Diagrams, Naming Conventions, and Design Issues 221
7.8 Example of Other Notation: UML Class Diagrams 226
7.9 Relationship Types of Degree Higher than Two 228
7.10 Summary 232
Review Questions 234
Exercises 234
Laboratory Exercises 241
Selected Bibliography 243

chapter 8 **The Enhanced Entity-Relationship (EER) Model 245**

8.1 Subclasses, Superclasses, and Inheritance 246
8.2 Specialization and Generalization 248
8.3 Constraints and Characteristics of Specialization and Generalization Hierarchies 251
8.4 Modeling of UNION Types Using Categories 258
8.5 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions 260
8.6 Example of Other Notation: Representing Specialization and Generalization in UML Class Diagrams 265
8.7 Data Abstraction, Knowledge Representation, and Ontology Concepts 267
8.8 Summary 273
Review Questions 273
Exercises 274
Laboratory Exercises 281
Selected Bibliography 284

chapter 9 **Relational Database Design by ER- and EER-to-Relational Mapping 285**

9.1 Relational Database Design Using ER-to-Relational Mapping 286
9.2 Mapping EER Model Constructs to Relations 294
9.3 Summary 299
Review Questions 299
Exercises 299
Laboratory Exercises 301
Selected Bibliography 302

chapter 10 **Practical Database Design Methodology and Use of UML Diagrams 303**

10.1 The Role of Information Systems in Organizations 304
10.2 The Database Design and Implementation Process 309
10.3 Use of UML Diagrams as an Aid to Database Design Specification 328
10.4 Rational Rose: A UML-Based Design Tool 337
10.5 Automated Database Design Tools 342

10.6 Summary 345
Review Questions 347
Selected Bibliography 348

part 4
Object, Object-Relational, and XML: Concepts, Models, Languages, and Standards

chapter 11 Object and Object-Relational Databases 353
11.1 Overview of Object Database Concepts 355
11.2 Object-Relational Features: Object Database Extensions to SQL 369
11.3 The ODMG Object Model and the Object Definition Language ODL 376
11.4 Object Database Conceptual Design 395
11.5 The Object Query Language OQL 398
11.6 Overview of the C++ Language Binding in the ODMG Standard 407
11.7 Summary 408
Review Questions 409
Exercises 411
Selected Bibliography 412

chapter 12 XML: Extensible Markup Language 415
12.1 Structured, Semistructured, and Unstructured Data 416
12.2 XML Hierarchical (Tree) Data Model 420
12.3 XML Documents, DTD, and XML Schema 423
12.4 Storing and Extracting XML Documents from Databases 431
12.5 XML Languages 432
12.6 Extracting XML Documents from Relational Databases 436
12.7 Summary 442
Review Questions 442
Exercises 443
Selected Bibliography 443

part 5
Database Programming Techniques

chapter 13 **Introduction to SQL Programming Techniques 447**

13.1 Database Programming: Techniques and Issues 448
13.2 Embedded SQL, Dynamic SQL, and SQLJ 451
13.3 Database Programming with Function Calls: SQL/CLI and JDBC 464
13.4 Database Stored Procedures and SQL/PSM 473
13.5 Comparing the Three Approaches 476
13.6 Summary 477
Review Questions 478
Exercises 478
Selected Bibliography 479

chapter 14 **Web Database Programming Using PHP 481**

14.1 A Simple PHP Example 482
14.2 Overview of Basic Features of PHP 484
14.3 Overview of PHP Database Programming 491
14.4 Summary 496
Review Questions 496
Exercises 497
Selected Bibliography 497

part 6
Database Design Theory and Normalization

chapter 15 **Basics of Functional Dependencies and Normalization for Relational Databases 501**

15.1 Informal Design Guidelines for Relation Schemas 503
15.2 Functional Dependencies 513
15.3 Normal Forms Based on Primary Keys 516
15.4 General Definitions of Second and Third Normal Forms 525
15.5 Boyce-Codd Normal Form 529

15.6 Multivalued Dependency and Fourth Normal Form 531
15.7 Join Dependencies and Fifth Normal Form 534
15.8 Summary 535
Review Questions 536
Exercises 537
Laboratory Exercises 542
Selected Bibliography 542

chapter 16 **Relational Database Design Algorithms and Further Dependencies 543**

16.1 Further Topics in Functional Dependencies: Inference Rules, Equivalence, and Minimal Cover 545
16.2 Properties of Relational Decompositions 551
16.3 Algorithms for Relational Database Schema Design 557
16.4 About Nulls, Dangling Tuples, and Alternative Relational Designs 563
16.5 Further Discussion of Multivalued Dependencies and 4NF 567
16.6 Other Dependencies and Normal Forms 571
16.7 Summary 575
Review Questions 576
Exercises 576
Laboratory Exercises 578
Selected Bibliography 579

part 7
File Structures, Indexing, and Hashing

chapter 17 **Disk Storage, Basic File Structures, and Hashing 583**

17.1 Introduction 584
17.2 Secondary Storage Devices 587
17.3 Buffering of Blocks 593
17.4 Placing File Records on Disk 594
17.5 Operations on Files 599
17.6 Files of Unordered Records (Heap Files) 601
17.7 Files of Ordered Records (Sorted Files) 603
17.8 Hashing Techniques 606

17.9 Other Primary File Organizations 616
17.10 Parallelizing Disk Access Using RAID Technology 617
17.11 New Storage Systems 621
17.12 Summary 624
Review Questions 625
Exercises 626
Selected Bibliography 630

chapter 18 Indexing Structures for Files 631

18.1 Types of Single-Level Ordered Indexes 632
18.2 Multilevel Indexes 643
18.3 Dynamic Multilevel Indexes Using B-Trees and B^+-Trees 646
18.4 Indexes on Multiple Keys 660
18.5 Other Types of Indexes 663
18.6 Some General Issues Concerning Indexing 668
18.7 Summary 670
Review Questions 671
Exercises 672
Selected Bibliography 674

part 8
Query Processing and Optimization, and Database Tuning

chapter 19 Algorithms for Query Processing and Optimization 679

19.1 Translating SQL Queries into Relational Algebra 681
19.2 Algorithms for External Sorting 682
19.3 Algorithms for SELECT and JOIN Operations 685
19.4 Algorithms for PROJECT and Set Operations 696
19.5 Implementing Aggregate Operations and OUTER JOINs 698
19.6 Combining Operations Using Pipelining 700
19.7 Using Heuristics in Query Optimization 700
19.8 Using Selectivity and Cost Estimates in Query Optimization 710
19.9 Overview of Query Optimization in Oracle 721
19.10 Semantic Query Optimization 722
19.11 Summary 723

Review Questions 723
Exercises 724
Selected Bibliography 725

chapter 20 Physical Database Design and Tuning 727

20.1 Physical Database Design in Relational Databases 727
20.2 An Overview of Database Tuning in Relational Systems 733
20.3 Summary 739
Review Questions 739
Selected Bibliography 740

part 9
Transaction Processing, Concurrency Control, and Recovery

chapter 21 Introduction to Transaction Processing Concepts and Theory 743

21.1 Introduction to Transaction Processing 744
21.2 Transaction and System Concepts 751
21.3 Desirable Properties of Transactions 754
21.4 Characterizing Schedules Based on Recoverability 755
21.5 Characterizing Schedules Based on Serializability 759
21.6 Transaction Support in SQL 770
21.7 Summary 772
Review Questions 772
Exercises 773
Selected Bibliography 775

chapter 22 Concurrency Control Techniques 777

22.1 Two-Phase Locking Techniques for Concurrency Control 778
22.2 Concurrency Control Based on Timestamp Ordering 788
22.3 Multiversion Concurrency Control Techniques 791
22.4 Validation (Optimistic) Concurrency Control Techniques 794
22.5 Granularity of Data Items and Multiple Granularity Locking 795
22.6 Using Locks for Concurrency Control in Indexes 798
22.7 Other Concurrency Control Issues 800

22.8 Summary 802
Review Questions 803
Exercises 804
Selected Bibliography 804

chapter 23 Database Recovery Techniques 807

23.1 Recovery Concepts 808
23.2 NO-UNDO/REDO Recovery Based on Deferred Update 815
23.3 Recovery Techniques Based on Immediate Update 817
23.4 Shadow Paging 820
23.5 The ARIES Recovery Algorithm 821
23.6 Recovery in Multidatabase Systems 825
23.7 Database Backup and Recovery from Catastrophic Failures 826
23.8 Summary 827
Review Questions 828
Exercises 829
Selected Bibliography 832

part 10
Additional Database Topics: Security and Distribution

chapter 24 Database Security 835

24.1 Introduction to Database Security Issues 836
24.2 Discretionary Access Control Based on Granting and Revoking Privileges 842
24.3 Mandatory Access Control and Role-Based Access Control for Multilevel Security 847
24.4 SQL Injection 855
24.5 Introduction to Statistical Database Security 859
24.6 Introduction to Flow Control 860
24.7 Encryption and Public Key Infrastructures 862
24.8 Privacy Issues and Preservation 866
24.9 Challenges of Database Security 867
24.10 Oracle Label-Based Security 868
24.11 Summary 870

Review Questions 872
Exercises 873
Selected Bibliography 874

chapter 25 Distributed Databases 877

25.1 Distributed Database Concepts 878
25.2 Types of Distributed Database Systems 883
25.3 Distributed Database Architectures 887
25.4 Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design 894
25.5 Query Processing and Optimization in Distributed Databases 901
25.6 Overview of Transaction Management in Distributed Databases 907
25.7 Overview of Concurrency Control and Recovery in Distributed Databases 909
25.8 Distributed Catalog Management 913
25.9 Current Trends in Distributed Databases 914
25.10 Distributed Databases in Oracle 915
25.11 Summary 919
Review Questions 921
Exercises 922
Selected Bibliography 924

part 11
Advanced Database Models, Systems, and Applications

chapter 26 Enhanced Data Models for Advanced Applications 931

26.1 Active Database Concepts and Triggers 933
26.2 Temporal Database Concepts 943
26.3 Spatial Database Concepts 957
26.4 Multimedia Database Concepts 965
26.5 Introduction to Deductive Databases 970
26.6 Summary 983
Review Questions 985
Exercises 986
Selected Bibliography 989

chapter 27 Introduction to Information Retrieval and Web Search 993

27.1 Information Retrieval (IR) Concepts 994
27.2 Retrieval Models 1001
27.3 Types of Queries in IR Systems 1007
27.4 Text Preprocessing 1009
27.5 Inverted Indexing 1012
27.6 Evaluation Measures of Search Relevance 1014
27.7 Web Search and Analysis 1018
27.8 Trends in Information Retrieval 1028
27.9 Summary 1030
Review Questions 1031
Selected Bibliography 1033

chapter 28 Data Mining Concepts 1035

28.1 Overview of Data Mining Technology 1036
28.2 Association Rules 1039
28.3 Classification 1051
28.4 Clustering 1054
28.5 Approaches to Other Data Mining Problems 1057
28.6 Applications of Data Mining 1060
28.7 Commercial Data Mining Tools 1060
28.8 Summary 1063
Review Questions 1063
Exercises 1064
Selected Bibliography 1065

chapter 29 Overview of Data Warehousing and OLAP 1067

29.1 Introduction, Definitions, and Terminology 1067
29.2 Characteristics of Data Warehouses 1069
29.3 Data Modeling for Data Warehouses 1070
29.4 Building a Data Warehouse 1075
29.5 Typical Functionality of a Data Warehouse 1078
29.6 Data Warehouse versus Views 1079
29.7 Difficulties of Implementing Data Warehouses 1080

29.8 Summary 1081
Review Questions 1081
Selected Bibliography 1082

appendix **A** **Alternative Diagrammatic Notations for ER Models 1083**

appendix **B** **Parameters of Disks 1087**

appendix **C** **Overview of the QBE Language 1091**

C.1 Basic Retrievals in QBE 1091
C.2 Grouping, Aggregation, and Database Modification in QBE 1095

appendix **D** **Overview of the Hierarchical Data Model**
(located on the Companion Website at http://www.aw.com/elmasri)

appendix **E** **Overview of the Network Data Model**
(located on the Companion Website at http://www.aw.com/elmasri)

Selected Bibliography 1099

Index 1133

This page intentionally left blank

part 1

Introduction to Databases