THE EXPERT'S VOICE® IN DATABASES

Database Systems

A Pragmatic Approach

—

A textbook introduction to database systems theory and practice

—

Second Edition

—

Elvis C. Foster
with Shripad Godbole

Apress®

Database Systems

A Pragmatic Approach

Second Edition

Elvis C. Foster
With Shripad Godbole

Apress®

Database Systems: A Pragmatic Approach

Elvis C. Foster
Associate Professor of Computer Science
Keene State College, New Hampshire, USA

Shripad Godbole
Senior Manager of Database Administration
Merkle Inc., Marlborough, Massachusetts, USA

ISBN-13 (pbk): 978-1-4842-1192-2 ISBN-13 (electronic): 978-1-4842-1191-5
DOI 10.1007/978-1-4842-1191-5

Library of Congress Control Number: 2016951944

Copyright © 2016 by Elvis C. Foster and Shripad Godbole

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director: Welmoed Spahr
Lead Editor: Jonathan Gennick
Development Editor: Douglas Pundick
Technical Reviewer: Louis Davidson and Michael Cunningham
Editorial Board: Steve Anglin, Pramila Balan, Laura Berendson, Aaron Black, Louise Corrigan, Jonathan Gennick, Robert Hutchinson, Celestin Suresh John, Nikhil Karkal, James Markham, Susan McDermott, Matthew Moodie, Natalie Pao, Gwenan Spearing
Coordinating Editor: Jill Balzano
Copy Editor: Mary Behr
Compositor: SPi Global
Indexer: SPi Global
Artist: SPi Global
Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springer.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this text is available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/.

Printed on acid-free paper

This book is dedicated to the honor of my late father, Claudius Foster,
who taught me the discipline of being a responsible, goal-driven person.
The book is also dedicated to my students — past, present, and future.
You are the object of my inspiration and motivation;
you are the reason for this work.

Contents at a Glance

About the Authors xxvii

About the Technical Reviewers xxix

Acknowledgments xxxi

Preface xxxiii

Part A: Preliminary Topics 1

Chapter 1: Introduction to Database Systems 3

Chapter 2: The Database System Environment 15

Part B: The Relational Database Model 35

Chapter 3: The Relational Model 37

Chapter 4: Integrity Rules and Normalization 73

Chapter 5: Database Modeling and Design 101

Chapter 6: Database User Interface Design 139

Chapter 7: Relational Algebra 155

Chapter 8: Relational Calculus 181

Chapter 9: Reflective Look at the Relational Database Model 195

Part C: The Structured Query Language 203

Chapter 10: Overview of SQL 205

Chapter 11: SQL Data Definition Statements 211

Chapter 12: SQL Data Manipulation Statements 249

Chapter 13: Logical Views and System Security 305

Chapter 14: The System Catalog 327
Chapter 15: Some Limitations of SQL 339

Part D: Advanced Topics 347
Chapter 16: Database Administration 349
Chapter 17: Distributed Database Systems 361
Chapter 18: Object Databases 371
Chapter 19: Data Warehousing 377
Chapter 20: Web-Accessible Databases 391
Chapter 21: Using Database Systems to Anchor Management Support Systems 403

Part E: Overview of Selected DBMS Suites 433
Chapter 22: Overview of Oracle 435
Chapter 23: Overview of DB2 443
Chapter 24: Overview of MySQL 451
Chapter 25: Overview of Microsoft SQL Server 461

Part F: Appendices 469
Appendix A1: Review of Trees 471
Appendix A2: Review of Hashing 505
Appendix A3: Review of Information-Gathering Techniques 523
Appendix A4: BNF Syntax for Selected SQL Statements 539
Appendix A5: Sample Exercises and Examination Questions 585

Index 597

Contents

About the Authors xxvii

About the Technical Reviewers xxix

Acknowledgments xxxi

Preface xxxiii

Part A: Preliminary Topics 1

Chapter 1: Introduction to Database Systems 3

1.1 Definitions and Rationale 3

1.2 Objectives of a Database System 6

1.2.1 Primary and Secondary Objectives 6

1.2.2 Clarification on Data Independence 7

1.3 Advantages of a Database System 8

1.4 Approaches to Database Design 8

1.4.1 Conventional Files 9

1.4.2 Database Approach 10

1.5 Desirable Features of a DBS 10

1.6 Database Development Life Cycle 11

1.7 Summary and Concluding Remarks 12

1.8 Review Questions 13

1.9 References and/or Recommended Readings 13

Chapter 2: The Database System Environment 15

2.1 Levels of Architecture 15

2.1.1 External Level 17

2.1.2 Conceptual Level 17

2.1.3 Internal Level 17

2.2 Inter-Level Mappings 18

2.3 Database Personnel 18

2.3.1 Database Administrator and Data Architect 18

2.3.2 Tools Expert 19

2.3.3 Application Programmer 19

2.3.4 User Liaison Specialist 20

2.3.5 Network and Infrastructure Specialist 20

2.4 The Database Management System 21

2.5 Components of DBMS Suite 23

2.5.1 The DBMS Engine 24

2.5.2 Data Definition Subsystem 24

2.5.3 The User Interface Subsystem 24

2.5.4 Application Development Subsystem 25

2.5.5 Data Administration Subsystem 25

2.5.6 Data Dictionary Subsystem 25

2.5.7 Data Communications Manager 25

2.5.8 Utilities Subsystem 26

2.6 Front-End and Back-End Perspectives 26

2.7 Database System Architecture 27

2.8 Database Management System Classifications 30

2.8.1 Classification Based on Data Models 30

2.8.2 Classification Based on Number of Users 30

2.8.3 Classification Based on Site Configuration 30

2.8.4 Classification Based on Database Purpose 31

2.9 Summary and Concluding Remarks 31
2.10 Review Questions 32
2.11 References and/or Recommended Readings 33

Part B: The Relational Database Model 35

Chapter 3: The Relational Model 37

3.1 Basic Concepts 37
3.2 Domains 40
3.3 Relations 42
3.3.1 Properties of a Relation 43
3.3.2 Kinds of Relations 43
3.4 Relational Database System 44
3.5 Identifying Relationships 44
3.6 Representing Relationships 46
3.6.1 The Entity-Relationship Model 46
3.6.2 The Object-Relationship Model 52
3.6.3 Summary of Traditional Models 52
3.6.4 Multiplicity of Relationships 54
3.7 Implementing Relationships 55
3.8 Innovation: Relation-Attributes List and Relationship List 59
3.9 Database Naming Convention 64
3.10 Summary and Concluding Remarks 69
3.11 Review Questions 70
3.12 References and/or Recommended Readings 71

Chapter 4: Integrity Rules and Normalization 73

4.1 Fundamental Integrity Rules 73
4.2 Foreign Key Concept 74
4.3 Rationale for Normalization 76

4.4 Functional Dependence and Non-Loss Decomposition 77
4.4.1 Functional Dependence 77
4.4.2 Non-Loss Decomposition 78
4.5 First Normal Form 81
4.5.1 Problems with Relations in 1NF Only 81
4.6 Second Normal Form 83
4.6.1 Problems with Relations in 2NF Only 83
4.7 Third Normal Form 84
4.7.1 Problems with Relations in 3NF Only 84
4.8 Boyce-Codd Normal Form 85
4.9 Fourth Normal Form 86
4.9.1 Multi-Valued Dependency 87
4.9.2 Fagin's Theorem 87
4.9.3 The Zoo Revisited 89
4.10 Fifth Normal Form 89
4.10.1 Definition of Join Dependency 91
4.10.2 Fagin's Theorem 91
4.11 An Example 92
4.12 Other Normal Forms 94
4.12.1 The Domain-Key Normal Form 94
4.12.2 The Sixth Normal Form 95
4.13 Summary and Concluding Remarks 97
4.14 Review Questions 99
4.15 References and/or Recommended Readings 100
Chapter 5: Database Modeling and Design 101
5.1 The Database Model and Database Design 101
5.1.1 The Database Model 102
5.1.2 Database Design 102
5.2 The E-R Model Revisited 102
5.3 Database Design via the E-R Model 106

5.4 The Extended Relational Model 106
5.4.1 Entity Classifications 106
5.4.2 Surrogates 107
5.4.3 E-Relations and P-Relations 109
5.4.4 Integrity Rules 112
5.5 Database Design via the XR Model 113
5.5.1 Determining the Kernel Entities 113
5.5.2 Determining the Characteristic Entities 113
5.5.3 Determining the Designative Entities 114
5.5.4 Determining the Associations 115
5.5.5 Determining Entity Subtypes and Super-types 116
5.5.6 Determining Component Entities 117
5.5.7 Determining Additional Properties 117
5.5.8 Additional Applications of the XR Model 119
5.6 The UML Model 119
5.7 Database Design via the UML Model 122
5.8 Innovation: The Object/Entity Specification Grid 123
5.9 Database Design via Normalization Theory 127
5.9.1 Example: Mountaineering Problem 127
5.9.2 Determining Candidate Keys and then Normalizing 130
5.10 Database Model and Design Tools 132
5.11 Summary and Concluding Remarks 134
5.12 Review Questions 135
5.13 References and/or Recommended Readings 136
Chapter 6: Database User Interface Design 139
6.1 Overview 139
6.2 Deciding on the User Interface 141
6.3 Steps in the User Interface Design 142
6.3.1 Menu or Graphical User Interface 142
6.3.2 Command-Based User Interface 145

6.4 User Interface Development and Implementation 145
6.5 Summary and Concluding Remarks 152
6.6 Review Questions 152
6.7 References and/or Recommend Readings 153

Chapter 7: Relational Algebra 155

7.1 Overview 155
7.2 Basic Operations of Relational Algebra 156
7.2.1 Primary and Secondary Operations 157
7.2.2 Codd's Original Classification of Operations 157
7.2.3 Nested Operations 157
7.3 Syntax of Relational Algebra 157
7.3.1 Select Statement 162
7.3.2 Projection Statement 163
7.3.3 Natural Join Statement 164
7.3.4 Cartesian Product 166
7.3.5 Theta-Join 167
7.3.6 Inner and Outer Joins 170
7.3.7 Union, Intersection, and Difference Statements 172
7.3.8 Division Statement 173
7.4 Aliases, Renaming, and the Relational Assignment 174
7.4.1 The Alias Operation 174
7.4.2 The Assignment Operation 175
7.4.3 The Rename Operation 175
7.5 Other Operators 176
7.6 Summary and Concluding Remarks 177
7.7 Review Questions 178
7.8 References and/or Recommended Readings 179

Chapter 8: Relational Calculus 181
8.1 Overview 181
8.2 Calculus Notations and Illustrations 183
8.3 Quantifiers, Free and Bound Variables 185
8.3.1 Well-Formed Formula 186
8.3.2 Free and Bound Variables 186
8.4 Substitution Rule and Standardization Rules 189
8.5 Introductory Query Optimization 189
8.6 Domain-Oriented Relational Calculus 192
8.7 Summary and Concluding Remarks 192
8.8 Review Questions 193
8.9 References and/or Recommended Readings 193
Chapter 9: Reflective Look at the Relational Database Model 195
9.1 The Relational Model Summarized 195
9.2 Ramifications of the Relational Model 196
9.2.1 Codd's Early Benchmark 196
9.2.2 Revised Definition of a Relational System 197
9.2.3 Far-Reaching Consequences 200
9.3 Summary and Concluding Remarks 200
9.4 Review Questions 200
9.5 References 201
Part C: The Structured Query Language 203
Chapter 10: Overview of SQL 205
10.1 Important Facts 205
10.1.1 Commonly Used DDL Statements 205
10.1.2 Commonly Used DML and DCL Statements 207
10.1.3 Syntax Convention 207

10.2 Advantages of SQL 207
10.3 Summary and Concluding Remarks 208
10.4 Review Questions 208
10.5 Recommended Readings 209

Chapter 11: SQL Data Definition Statements 211

11.1 Overview of Oracle's SQL Environment 212
11.2 Basic Concepts in a Typical Oracle Database Environment 213
11.3 Database Creation 215
11.4 Database Management 219
11.5 Tablespace Creation 223
11.6 Tablespace Management 226
11.7 Table Creation 228
11.8 Dropping or Modifying a Table 234
11.9 Working with Indexes 236
11.10 Working with Sequences 244
11.10.1 Creating and Using Sequences 244
11.10.2 Altering and Dropping Sequences 245
11.11 Working with Synonyms 246
11.12 Summary and Concluding Remarks 247
11.13 Review Questions 247
11.14 References and/or Recommended Readings 248

Chapter 12: SQL Data Manipulation Statements 249

12.1 Insertion of Data 249
12.2 Update Operations 252
12.3 Deletion of Data 255
12.4 Commit and Rollback Operations 256
12.5 Basic Syntax for Queries 258
12.6 Simple Queries 261

12.7 Queries Involving Multiple Tables 263
12.7.1 The Traditional Method 263
12.7.2 The ANSI Method 266
12.8 Queries Involving the Use of Functions 267
12.8.1 Row Functions 268
12.8.2 Date Functions 270
12.8.3 Data Conversion Functions 272
12.8.4 Programmer-Defined Functions 274
12.8.5 Aggregation Functions 275
12.8.6 Analytic Functions 279
12.9 Queries Using the LIKE and BETWEEN Operators 280
12.10 Nested Queries 281
12.10.1 Nested Queries Involving Use of the IN Operator 282
12.10.2 Nested Queries Involving use of ANY or ALL Operator 283
12.10.3 Nested Queries Using the Existential Quantifier 285
12.11 Queries Involving Set Operations 285
12.12 Queries with Runtime Variables 287
12.13 Queries Involving SQL*Plus Format Commands 288
12.14 Embedded SQL 289
12.15 Dynamic Queries 295
12.16 Summary and Concluding Remarks 301
12.17 Review Questions 303
12.18 References and/or Recommended Readings 303
Chapter 13: Logical Views and System Security 305
13.1 Traditional Logical Views 305
13.1.1 View Creation 306
13.1.2 View Modification and Removal 308
13.1.3 Usefulness and Manipulation of Logical Views 309

13.2 System Security 309
13.2.1 Access to the System 310
13.2.2 Access to the System Resources 314
13.2.3 Access to the System Data 317
13.3 Materialized Views 320
13.3.1 Creating a Materialized View 320
13.3.2 Altering or Dropping a Materialized View 322
13.4 Summary and Concluding Remarks 324
13.5 Review Questions 324
13.6 References and/or Recommended Readings 325
Chapter 14: The System Catalog 327
14.1 Introduction 327
14.2 Three Important Catalog Tables 328
14.2.1 The User_Tables View 328
14.2.2 The User_Tab_Columns View 329
14.2.3 The User_Indexes View 329
14.3 Other Important Catalog Tables 331
14.4 Querying the System Catalog 334
14.5 Updating the System Catalog 336
14.6 Summary and Concluding Remarks 337
14.7 Review Questions 338
14.8 References and/or Recommended Readings 338
Chapter 15: Some Limitations of SQL 339
15.1 Programming Limitations 339
15.2 Limitations on Views 339
15.2.1 Restriction on Use of the Order-By Clause for Earlier Versions of SQL 340
15.2.2 Restriction on Data Manipulation for Views involving UNION, INTERSECT, or JOIN 340
15.2.3 Restriction on the Use of Aggregation Functions for Earlier Versions of SQL 341

15.3 Stringent Enforcement of Referential Integrity 341
15.4 Limitations on Calculated Columns 343
15.5 If-Then Limitation 344
15.6 Summary and Concluding Remarks 344
15.7 Review Questions 345
15.8 Recommended Readings 345

Part D: Advanced Topics 347

Chapter 16: Database Administration 349

16.1 Database Installation, Creation, and Configuration 349
16.2 Database Security 350
16.3 Database Management 350
16.4 Database Backup and Recovery 351
16.4.1 Oracle Backups: Basic Concept 351
16.4.2 Oracle Recovery: Basic Concept 352
16.4.3 Types of Failures 352
16.4.4 Database Backups 353
16.4.5 Basic Recovery Steps 354
16.4.6 Oracle's Backup and Recovery Solutions 354
16.5 Database Tuning 355
16.5.1 Tuning Goals 355
16.5.2 Tuning Methodology 356
16.6 Database Removal 358
16.7 Summary and Concluding Remarks 358
16.8 Review Questions 359
16.9 References and/or Recommended Readings 359

Chapter 17: Distributed Database Systems 361

17.1 Preliminaries 361
17.2 Advantages of Distributed Database Systems 362
17.3 12 Rules for Distributed Database Systems 363

17.4 Challenges to Distributed Database Systems 365
17.4.1 Query Optimization 365
17.4.2 Catalog Management 365
17.4.3 Update Propagation 366
17.4.4 Concurrency 366
17.4.5 Transaction Management 367
17.5 Database Gateways 367
17.6 The Future of Distributed Database Systems 368
17.6.1 Object Technology 368
17.6.2 Electronic Communication Systems 368
17.6.3 Cloud Technology 369
17.6.4 Big Data 369
17.7 Summary and Concluding Remarks 369
17.8 Review Questions 370
17.9 References and/or Recommended Readings 370
Chapter 18: Object Databases 371
18.1 Overview 371
18.2 Overview of Object-Oriented Database Management Systems 373
18.3 Challenges for Object-Oriented Database Management Systems 374
18.4 Hybrid Approach 375
18.5 Summary and Concluding Remarks 375
18.6 Review Questions 376
18.7 References and/or Recommended Readings 376
Chapter 19: Data Warehousing 377
19.1 Overview 377
19.1.1 Definitions 378
19.1.2 Acquiring a Data Warehouse 378

19.2 Rationale for Data Warehousing 379
19.3 Characteristics of a Data Warehouse 380
19.3.1 Definitive Features 380
19.3.2 Nature of Data Stored 380
19.3.3 Processing Requirements 381
19.3.4 12 Rules for Data Warehousing 383
19.4 Data Warehouse Architecture 384
19.4.1 Basic Data Warehouse Architecture 384
19.4.2 Data Warehouse Architecture with a Staging Area 384
19.4.3 Data Warehouse Architecture with a Staging Area and Data Marts 385
19.5 Extraction, Transformation, and Loading 387
19.5.1 What Happens During the ETL Process 387
19.5.2 ETL Tools 387
19.5.3 Daily Operations and Expansion of the Data Warehouse 388
19.6 Summary and Concluding Remarks 388
19.7 Review Questions 389
19.8 References and/or Recommended Readings 390
Chapter 20: Web-Accessible Databases 391
20.1 Introduction 391
20.2 Web-Accessible Database Architecture 392
20.3 Supporting Technologies 393
20.4 Implementation with Selected DBMS Suites 395
20.4.1 Implementation via Oracle 395
20.4.2 Implementation via DB2 396
20.4.3 Implementation via MySQL 397
20.5 Generic Implementation via Front-End and Back-End Tools 397
20.6 Challenges and Opportunities 398
20.6.1 Cloud Computing and Big Data 398
20.6.2 Cybersecurity 399

20.7 Summary and Concluding Remarks 400
20.8 Review Questions 400
20.9 References and/or Recommended Readings 401

Chapter 21: Using Database Systems to Anchor Management Support Systems 403

21.1 Overview of Management Support Systems 403
21.2 Building System Security Through Database Design 404
21.3 Case Study: Dynamic Menu Interface Designer 405
21.3.1 Database Requirements of the DMID 406
21.3.2 Overview of the DMID's User Interface Requirements 412
21.3.3 Management of System Constraints via the DMID 413
21.3.4 Access to System Resources 415
21.3.5 Usefulness and Applicability of the DMID 419
21.4 Selected MSS Project Ideas 419
21.4.1 Electoral Management System 419
21.4.2 Health Information Management System 420
21.4.3 Strategic Education Management System 421
21.4.4 Flight Information Management System 422
21.4.5 Financial Information Management System 422
21.4.6 Disease Recognition System 423
21.4.7 Cognitive Leadership Analysis System 424
21.4.8 Financial Status Assessment System 426
21.4.9 College/University Administrative Information System 427
21.5 Summary and Concluding Remarks 429
21.6 Review Questions 431
21.7 References and/or Recommended Readings 431

Part E: Overview of Selected DBMS Suites 433

Chapter 22: Overview of Oracle 435

22.1 Introduction 435

22.2 Main Components of the Oracle Suite 437

22.2.1 Oracle Server 437

22.2.2 Oracle PL/SQL and SQL*Plus 438

22.2.3 Oracle Developer Suite 438

22.2.4 Oracle Enterprise Manager Database Control and SQL Developer 439

22.2.5 Oracle Enterprise Manager Grid Control 439

22.2.6 Oracle Database Configuration Assistant 439

22.2.7 Oracle Warehouse Builder 439

22.3 Shortcomings of Oracle 440

22.4 Summary and Concluding Remarks 441

22.5 Review Questions 441

22.6 References and/or Recommended Readings 442

Chapter 23: Overview of DB2 443

23.1 Introduction 443

23.2 Main Components of the DB2 Suite 445

23.2.1 DB2 Universal Database Core 446

23.2.2 Command Line Processor Plus (CLPPlus) 447

23.2.3 IBM InfoSphere Information Server 447

23.2.4 IBM Data Studio 448

23.2.5 IBM InfoSphere Warehouse 449

23.3 Shortcomings of DB2 449

23.4 Summary and Concluding Remarks 450

23.5 Review Questions 450

23.6 References and/or Recommended Readings 450

Chapter 24: Overview of MySQL 451

24.1 Introduction to MySQL 451

24.2 Main Features of MySQL 454

24.3 Main Components of MySQL 456

24.4 Alternate Storage Engines 457

24.5 Shortcomings of MySQL 457

24.5.1 Limitations on Logical Views 458

24.5.2 Limitations on Subqueries 458

24.5.3 Limitation on Server-Side Cursors 459

24.5.4 Other Limitations 459

24.6 Summary and Concluding Remarks 460

24.7 Review Questions 460

24.8 References and/or Recommended Readings 460

Chapter 25: Overview of Microsoft SQL Server 461

25.1 Introduction 461

25.1.1 Brief History 461

25.1.2 Operating Environment 462

25.1.3 Microsoft SQL Server Editions 462

25.2 Main Features of Microsoft SQL Server 462

25.3 Main Components of Microsoft SQL Server Suite 463

25.3.1 Server Components 463

25.3.2 Management Tools 464

25.3.3 Development Tools 464

25.3.4 Client Connectivity 465

25.3.5 Code Samples 465

25.4 Shortcomings of Microsoft SQL Server 466

25.5 Summary and Concluding Remarks 467

25.6 Review Questions 467

25.7 References and/or Recommended Readings 467

Part F: Appendices **469**

Appendix A1: Review of Trees **471**

A1.1 Introduction to Trees 471

A1.2 Binary Trees 472

A1.2.1 Overview of Binary Trees 472

A1.2.2 Representation of Binary Trees 475

A1.2.3 Application of Binary Trees 476

A1.2.4 Operations on Binary Trees 476

A1.2.5 Implementation of Binary Trees 476

A1.2.6 Binary Tree Traversals 482

A1.2.7 Using Binary Trees to Evaluate Expressions 485

A1.3 Threaded Binary Trees 487

A1.4 Binary Search Trees 488

A1.5 Height-Balanced Trees 491

A1.6 Heaps 492

A1.6.1 Building the Heap 492

A1.6.2 Processing the Heap (Heap-Sort) 493

A1.7 M-Way Search Trees and B-Trees 495

A1.7.1 Definition of B-Tree 497

A1.7.2 Implementation of the B-tree 499

A1.8 Summary and Concluding Remarks 502

A1.9 References and/or Recommended Readings 503

Appendix A2: Review of Hashing **505**

A2.1 Introduction 505

A2.2 Hash Functions 506

A2.2.1 Absolute Addressing 506

A2.2.2 Direct Table Lookup 506

A2.2.3 Division-Remainder 508

A2.2.4 Mid-Square 508

A2.2.5 Folding 509

A2.2.6 Truncation 509
A2.2.7 Treating Alphanumeric Key Values 510
A2.3 Collision Resolution 510
A2.3.1 Linear Probing 510
A2.3.2 Synonym Chaining 511
A2.3.3 Rehashing 513
A2.4 Hashing in Java 513
A2.5 Summary and Concluding Remarks 522
A2.6 References and/or Recommended Readings 522
Appendix A3: Review of Information-Gathering Techniques 523
A3.1 Rationale for Information Gathering 523
A3.2 Interviewing 525
A3.2.1 Steps in Planning the Interview 525
A3.2.2 Basic Guidelines for Interviews 525
A3.3 Questionnaires and Surveys 527
A3.3.1 Guidelines for Questionnaires 527
A3.3.2 Using Scales in Questionnaires 527
A3.3.3 Administering the Questionnaire 528
A3.4 Sampling 528
A3.4.1 Probability Sampling Techniques 529
A3.4.2 Non-Probability Sampling Techniques 529
A3.4.3 Sample Calculations 529
A3.5 Observation and Document Review 532
A3.6 Prototyping 532
A3.7 Brainstorming 533
A3.8 Object Identification 533
A3.8.1 The Descriptive Narrative Approach 534
A3.8.2 The Rule-of-Thumb Approach 535
A3.9 Summary and Concluding Remarks 537
A3.10 References and/or Recommended Readings 538

Appendix A4: BNF Syntax for Selected SQL Statements 539

A4.1 Database Management 540

A4.1.1 Syntax for **Create-Database** Statement 540

A4.1.2 Syntax for **Alter-Database** Statement 541

A4.2 Tablespace Management 545

A4.2.1 Syntax for **Create-Tablespace** Statement 545

A4.2.2 Syntax for **Alter-Tablespace** Statement 546

A4.3 Tables Management 547

A4.3.1 Syntax for **Create-Table** Statement 547

A4.3.2 Syntax for **Alter-Table** Statement 555

A4.4 Index Management 562

A4.4.1 Syntax for **Create-Index** Statement 562

A4.4.2 Syntax for **Alter-Index** Statement 564

A4.5 Data Insertion, Update, and Deletion 566

A4.5.1 Abridged Syntax for the **Insert** Statement 566

A4.5.2 Abridged Syntax for the **Update** Statement 566

A4.5.3 Syntax for **Delete** Statement and **Truncate** Statement 567

A4.6 Transaction Management 567

A4.7 Building Queries 568

A4.7.1 Abridged Syntax for the **Select** Statement 568

A4.7.2 Modified **From-Clause** for ANSI Join 569

A4.7.3 Some Commonly Used Row Functions 570

A4.7.4 Some Commonly Used Date Manipulation Functions 573

A4.7.5 Some Commonly Used Data Conversion Functions 574

A4.7.6 Some Valid Date and Numeric Formats 575

A4.7.7 Commonly Used Aggregation Functions 576

A4.7.8 Syntax for Using Analytic Functions 577

A4.7.9 Syntax for Nested Queries 577

A4.8 Managing Logical Views 578

A4.8.1 Creating the Logical View 578

A4.8.2 Altering or Dropping the Logical View 579

A4.9 Managing System Security 580
A4.9.1 Syntax for **Create-Profile** Statement 580
A4.9.2 Syntax for Altering or Dropping a Profile 581
A4.9.3 Syntax for Creating User Account(s) 581
A4.9.4 Syntax for Altering or Dropping User Account(s) 582
A4.9.5 Syntax for Creating, Altering, or Dropping Role(s) 582
A4.9.6 Syntax for Granting or Revoking Privilege(s) 583

Appendix A5: Sample Exercises and Examination Questions 585
A5.1 Introduction 585
A5.2 Sample Assignment 2A 586
A5.3 Sample Assignment 3B 586
A5.4 Sample Assignment 4A 588
A5.5 Sample Assignment 5A 590
A5.6 Sample Assignment 6A 591
A6.7 Sample Assignment 7A 592
A5.8 Sample Assignment 8 593
A5.9 Sample Assignment 9 594

Index 597

About the Authors

Elvis C. Foster is Associate Professor of Computer Science at Keene State College, New Hampshire. He holds a Bachelor of Science (BSc) in Computer Science and Electronics, as well as a Doctor of Philosophy (PhD) in Computer Science (specializing in strategic information systems and database systems) from University of the West Indies, Mona Jamaica. Dr. Foster has over 25 years of combined experience as a software engineer, database expert, information technology executive and consultant, and computer science educator. He has had the favorable experience of being involved with the design and/or development of software systems for various medium-sized and large organizations, including the central bank of his own country. He has lectured at the tertiary level in three different countries, including the United States, where he now resides. He has produced many outstanding computer science and information technology professionals. Many of them have excelled at graduate school; others have gone straight to industry after their first degree; the thorough grounding acquired often serves them well, be it in leading software engineering enterprises around the world, or in other chosen vocations. His textbooks draw from that experience.

Shripad V. Godbole is an independent database administrator/consultant with over 20 years of experience in diverse business environments, information infrastructure planning, diagnostics, and administration. His qualifications include Bachelor of Science (BS) in Physics, Bachelor of Computer Science (BCS), Master of Science (MS) in Physics with specialization in Electronics, all from Poona University in Pune, India. He is also an Oracle Certified Professional Database Administrator (OCPDBA), and holds a Master of Business Administration (MBA) in Technology Management from University of Phoenix.

About the Technical Reviewers

Louis Davidson has been in the IT industry for more than 15 years as a corporate database developer and architect. He has spent the majority of his career working with Microsoft SQL Server, beginning from the early days of version 1.0. He has a bachelor's degree from the University of Tennessee at Chattanooga in Computer Science, with a minor in Mathematics. Louis is the data architect for Compass Technology (Compass.net) in Chesapeake, Virginia, leading database development on their suite of nonprofit-oriented CRM products, built on the Microsoft CRM platform and SQL Server technologies.

Michael Cunningham has been in IT for 23 years. The first 10 years he spent as a software engineer, and then converted to being an Oracle DBA. I suppose you could say he's a real Oracle geek since he totally enjoys it. There is always something to learn and share.
In his spare time, he enjoys sailing, archery, RVing, and a bunch of other things that keep him away from being "connected"
(because we all need a break from time to time).

Acknowledgments

I express profound gratitude to my wife, Jacqueline, and to my children, Chris-Ann and K. V. Rhoden, for putting up with me during the extended periods of preparation of this text. There have been countless nights spent in my study at home, and as many forgone opportunities for family time; still my family members understood that this project meant a lot to me, and their support was unwavering.

Also, I must recognize several of my past and current students at Keene State College (KSC), on whom my lecture notes have been repeatedly tested. Of the groups taught at the institution, I must single out the contributions of a few students: Georgie Hill committed personal time to assist me in redoing many of the illustrations; Jesse Schmidt, James Dahlen, Joshua Pritchett, Myles Dumas, and Thomas O'Dea participated on various ongoing projects, some of which have been mentioned in Chapter 21.

Prior to my stint at KSC, I had the good fortune of grooming several students from three different institutions. These earlier experiences were foundational for the eventual publication of this volume. In this regard, I would like to make special mention of Dionne Jackson, Kerron Hislop, Brigid Winter, Abrams O'Buyonge, Sheldon Kennedy, and Ruth Del Rosario.

Special appreciation is offered to my colleague Shripad Godbole, who has coauthored some of the chapters with me, particularly in Division E of the text. Being a practicing database administrator, Shripad also served as a valuable resource in these areas.

The editorial and production teams at Apress Publishing deserve mention for their work in facilitating the publication of this volume. I make special mention of Jonathan Gennick, Jill Balzano, and Douglas Pundick. Finally, authors Louis Davidson and Michael Cunningham fulfilled the role of peer reviewers of the manuscript; they were exceptional, and I thank them. Earlier editions and drafts were reviewed by Han Reichgelt, Marlon Moncrieffe, and Jacob Mangal; they also deserve mention. Thanks to everyone!

—Elvis C. Foster, PhD; Keene State College; Keene, New Hampshire, USA

Preface

This book has been compiled with three target groups in mind. The book is best suited for undergraduate students of computer science (CS) or a related discipline who are pursuing a course in database systems. Graduate students who are pursuing an introductory course in the subject may also find it useful. Finally, practicing software engineers and/or information technology (IT) professionals who need a quick reference on database design may find it useful.

The motivation that drove this work was a desire to provide a concise but comprehensive guide to the discipline of database design, construction, implementation, and management. Having worked in the software engineering and IT industries for several years before making a career switch to academia, it has been my observation that many IT professionals and software engineers tend to pay little attention to their database design skills; this is often reflected in the proliferation of software applications with inadequately designed underlying databases. In this text, the discipline of database systems design and management is discussed within the context of a bigger picture: that of software engineering. The reader is led to understand from the outset that a database is a critical component of a software system, and that proper database design and management is integral to the success of the software system. Additionally and simultaneously, the reader is led to appreciate the huge value of a properly designed database to the success of a business enterprise.

The book draws from lecture notes that have been compiled and tested over several years, and with outstanding results. The lecture notes draw on personal experiences gained in the industry over the years, as well as the suggestions of various professionals and students. The chapters are organized in a manner that reflects my own approach in teaching the course, but each chapter may be read on its own.

The text has been prepared specifically to meet three objectives: comprehensive coverage, brevity, and relevance. Comprehensive coverage and brevity often operate as competing goals. In order to achieve both, I have adopted a pragmatic approach that gets straight to the critical issues for each topic, and avoids unnecessary fluff, while using the question of relevance as the balancing force. Additionally, readers should find the following features quite convenient:

- Short paragraphs that express the salient aspects of the subject matter being discussed
- Bullet points or numbers to itemize important things to be remembered
- Diagrams and illustrations to enhance the reader's understanding
- Real-to-life examples
- Introduction of a few original methodologies that are applicable to database design. The main ones are the Relations-Attributes List (RAL, as in section 3.8); the object naming convention (ONC, as in section 3.9); the object/entity specification grid (O/ESG, as in section 5.8); the user interface topology chart (UITC, as in section 6.3); the dynamic menu interface designer (DMID, as in section 21.3); and the information topology chart (ITC, as in subsection 21.4.9).
- Step-by-step, reader-friendly guidelines for solving generic database systems problems