Contents

Chapter 1 Introduction

1.1 Database-System Applications 1
1.2 Purpose of Database Systems 3
1.3 View of Data 6
1.4 Database Languages 9
1.5 Relational Databases 12
1.6 Database Design 15
1.7 Data Storage and Querying 20
1.8 Transaction Management 22
1.9 Database Architecture 23
1.10 Data Mining and Information Retrieval 25
1.11 Specialty Databases 26
1.12 Database Users and Administrators 27
1.13 History of Database Systems 29
1.14 Summary 31
Exercises 33
Bibliographical Notes 35

PART ONE ■ RELATIONAL DATABASES

Chapter 2 Introduction to the Relational Model

2.1 Structure of Relational Databases 39
2.2 Database Schema 42
2.3 Keys 45
2.4 Schema Diagrams 46
2.5 Relational Query Languages 47
2.6 Relational Operations 48
2.7 Summary 52
Exercises 53
Bibliographical Notes 55

Chapter 3 Introduction to SQL

3.1 Overview of the SQL Query Language 57
3.2 SQL Data Definition 58
3.3 Basic Structure of SQL Queries 63
3.4 Additional Basic Operations 74
3.5 Set Operations 79
3.6 Null Values 83
3.7 Aggregate Functions 84
3.8 Nested Subqueries 90
3.9 Modification of the Database 98
3.10 Summary 104
Exercises 105
Bibliographical Notes 112

Chapter 4 Intermediate SQL

4.1 Join Expressions 113
4.2 Views 120
4.3 Transactions 127
4.4 Integrity Constraints 128
4.5 SQL Data Types and Schemas 136
4.6 Authorization 143
4.7 Summary 150
Exercises 152
Bibliographical Notes 156

Chapter 5 Advanced SQL

5.1 Accessing SQL From a Programming Language 157
5.2 Functions and Procedures 173
5.3 Triggers 180
5.4 Recursive Queries** 187
5.5 Advanced Aggregation Features** 192
5.6 OLAP** 197
5.7 Summary 209
Exercises 211
Bibliographical Notes 216

Chapter 6 Formal Relational Query Languages

6.1 The Relational Algebra 217
6.2 The Tuple Relational Calculus 239
6.3 The Domain Relational Calculus 245
6.4 Summary 248
Exercises 249
Bibliographical Notes 254

PART TWO ■ DATABASE DESIGN

Chapter 7 Database Design and the E-R Model

7.1 Overview of the Design Process 259
7.2 The Entity-Relationship Model 262
7.3 Constraints 269
7.4 Removing Redundant Attributes in Entity Sets 272
7.5 Entity-Relationship Diagrams 274
7.6 Reduction to Relational Schemas 283
7.7 Entity-Relationship Design Issues 290
7.8 Extended E-R Features 295
7.9 Alternative Notations for Modeling Data 304
7.10 Other Aspects of Database Design 310
7.11 Summary 313
Exercises 315
Bibliographical Notes 321

Chapter 8 Relational Database Design

8.1 Features of Good Relational Designs 323
8.2 Atomic Domains and First Normal Form 327
8.3 Decomposition Using Functional Dependencies 329
8.4 Functional-Dependency Theory 338
8.5 Algorithms for Decomposition 348
8.6 Decomposition Using Multivalued Dependencies 355
8.7 More Normal Forms 360
8.8 Database-Design Process 361
8.9 Modeling Temporal Data 364
8.10 Summary 367
Exercises 368
Bibliographical Notes 374

Chapter 9 Application Design and Development

9.1 Application Programs and User Interfaces 375
9.2 Web Fundamentals 377
9.3 Servlets and JSP 383
9.4 Application Architectures 391
9.5 Rapid Application Development 396
9.6 Application Performance 400
9.7 Application Security 402
9.8 Encryption and Its Applications 411
9.9 Summary 417
Exercises 419
Bibliographical Notes 426

PART THREE ■ DATA STORAGE AND QUERYING

Chapter 10 Storage and File Structure

10.1 Overview of Physical Storage Media 429
10.2 Magnetic Disk and Flash Storage 432
10.3 RAID 441
10.4 Tertiary Storage 449
10.5 File Organization 451
10.6 Organization of Records in Files 457
10.7 Data-Dictionary Storage 462
10.8 Database Buffer 464
10.9 Summary 468
Exercises 470
Bibliographical Notes 473

Chapter 11 Indexing and Hashing

11.1 Basic Concepts 475
11.2 Ordered Indices 476
11.3 B^+-Tree Index Files 485
11.4 B^+-Tree Extensions 500
11.5 Multiple-Key Access 506
11.6 Static Hashing 509
11.7 Dynamic Hashing 515
11.8 Comparison of Ordered Indexing and Hashing 523
11.9 Bitmap Indices 524
11.10 Index Definition in SQL 528
11.11 Summary 529
Exercises 532
Bibliographical Notes 536

Chapter 12 Query Processing

12.1 Overview 537
12.2 Measures of Query Cost 540
12.3 Selection Operation 541
12.4 Sorting 546
12.5 Join Operation 549
12.6 Other Operations 563
12.7 Evaluation of Expressions 567
12.8 Summary 572
Exercises 574
Bibliographical Notes 577

Chapter 13 Query Optimization

13.1 Overview 579
13.2 Transformation of Relational Expressions 582
13.3 Estimating Statistics of Expression Results 590
13.4 Choice of Evaluation Plans 598
13.5 Materialized Views** 607
13.6 Advanced Topics in Query Optimization** 612
13.7 Summary 615
Exercises 617
Bibliographical Notes 622

PART FOUR ■ TRANSACTION MANAGEMENT

Chapter 14 Transactions

14.1 Transaction Concept 627
14.2 A Simple Transaction Model 629
14.3 Storage Structure 632
14.4 Transaction Atomicity and Durability 633
14.5 Transaction Isolation 635
14.6 Serializability 640
14.7 Transaction Isolation and Atomicity 647
14.8 Transaction Isolation Levels 648
14.9 Implementation of Isolation Levels 650
14.10 Transactions as SQL Statements 653
14.11 Summary 655
Exercises 657
Bibliographical Notes 660

Chapter 15 Concurrency Control

15.1 Lock-Based Protocols 661
15.2 Deadlock Handling 674
15.3 Multiple Granularity 679
15.4 Timestamp-Based Protocols 682
15.5 Validation-Based Protocols 686
15.6 Multiversion Schemes 689
15.7 Snapshot Isolation 692
15.8 Insert Operations, Delete Operations, and Predicate Reads 697
15.9 Weak Levels of Consistency in Practice 701
15.10 Concurrency in Index Structures** 704
15.11 Summary 708
Exercises 712
Bibliographical Notes 718

Chapter 16 Recovery System

16.1 Failure Classification 721
16.2 Storage 722
16.3 Recovery and Atomicity 726
16.4 Recovery Algorithm 735
16.5 Buffer Management 738
16.6 Failure with Loss of Nonvolatile Storage 743
16.7 Early Lock Release and Logical Undo Operations 744
16.8 ARIES** 750
16.9 Remote Backup Systems 756
16.10 Summary 759
Exercises 762
Bibliographical Notes 766

PART FIVE ■ SYSTEM ARCHITECTURE

Chapter 17 Database-System Architectures

17.1 Centralized and Client–Server Architectures 769
17.2 Server System Architectures 772
17.3 Parallel Systems 777
17.4 Distributed Systems 784
17.5 Network Types 788
17.6 Summary 791
Exercises 793
Bibliographical Notes 794

Chapter 18 Parallel Databases

18.1 Introduction 797
18.2 I/O Parallelism 798
18.3 Interquery Parallelism 802
18.4 Intraquery Parallelism 803
18.5 Intraoperation Parallelism 804
18.6 Interoperation Parallelism 813
18.7 Query Optimization 814
18.8 Design of Parallel Systems 815
18.9 Parallelism on Multicore Processors 817
18.10 Summary 819
Exercises 821
Bibliographical Notes 824

Chapter 19 Distributed Databases

19.1 Homogeneous and Heterogeneous Databases 825
19.2 Distributed Data Storage 826
19.3 Distributed Transactions 830
19.4 Commit Protocols 832
19.5 Concurrency Control in Distributed Databases 839
19.6 Availability 847
19.7 Distributed Query Processing 854
19.8 Heterogeneous Distributed Databases 857
19.9 Cloud-Based Databases 861
19.10 Directory Systems 870
19.11 Summary 875
Exercises 879
Bibliographical Notes 883

PART SIX ■ DATA WAREHOUSING, DATA MINING, AND INFORMATION RETRIEVAL

Chapter 20 Data Warehousing and Mining

20.1 Decision-Support Systems 887
20.2 Data Warehousing 889
20.3 Data Mining 893
20.4 Classification 894
20.5 Association Rules 904
20.6 Other Types of Associations 906
20.7 Clustering 907
20.8 Other Forms of Data Mining 908
20.9 Summary 909
Exercises 911
Bibliographical Notes 914

Chapter 21 Information Retrieval

21.1 Overview 915
21.2 Relevance Ranking Using Terms 917
21.3 Relevance Using Hyperlinks 920
21.4 Synonyms, Homonyms, and Ontologies 925
21.5 Indexing of Documents 927
21.6 Measuring Retrieval Effectiveness 929
21.7 Crawling and Indexing the Web 930
21.8 Information Retrieval: Beyond Ranking of Pages 931
21.9 Directories and Categories 935
21.10 Summary 937
Exercises 939
Bibliographical Notes 941

PART SEVEN ■ SPECIALTY DATABASES

Chapter 22 Object-Based Databases

22.1 Overview 945
22.2 Complex Data Types 946
22.3 Structured Types and Inheritance in SQL 949
22.4 Table Inheritance 954
22.5 Array and Multiset Types in SQL 956
22.6 Object-Identity and Reference Types in SQL 961
22.7 Implementing O-R Features 963
22.8 Persistent Programming Languages 964
22.9 Object-Relational Mapping 973
22.10 Object-Oriented versus Object-Relational 973
22.11 Summary 975
Exercises 976
Bibliographical Notes 980

Chapter 23 XML

23.1 Motivation 981
23.2 Structure of XML Data 986
23.3 XML Document Schema 990
23.4 Querying and Transformation 998
23.5 Application Program Interfaces to XML 1008
23.6 Storage of XML Data 1009
23.7 XML Applications 1016
23.8 Summary 1019
Exercises 1021
Bibliographical Notes 1024

PART EIGHT ■ ADVANCED TOPICS

Chapter 24 Advanced Application Development

24.1 Performance Tuning 1029
24.2 Performance Benchmarks 1045
24.3 Other Issues in Application Development 1048
24.4 Standardization 1051
24.5 Summary 1056
Exercises 1057
Bibliographical Notes 1059

Chapter 25 Spatial and Temporal Data and Mobility

25.1 Motivation 1061
25.2 Time in Databases 1062
25.3 Spatial and Geographic Data 1064
25.4 Multimedia Databases 1076
25.5 Mobility and Personal Databases 1079
25.6 Summary 1085
Exercises 1087
Bibliographical Notes 1089

Chapter 26 Advanced Transaction Processing

26.1 Transaction-Processing Monitors 1091
26.2 Transactional Workflows 1096
26.3 E-Commerce 1102
26.4 Main-Memory Databases 1105
26.5 Real-Time Transaction Systems 1108
26.6 Long-Duration Transactions 1109
26.7 Summary 1115
Exercises 1117
Bibliographical Notes 1119

PART NINE ■ CASE STUDIES

Chapter 27 PostgreSQL

27.1 Introduction 1123
27.2 User Interfaces 1124
27.3 SQL Variations and Extensions 1126
27.4 Transaction Management in PostgreSQL 1137
27.5 Storage and Indexing 1146
27.6 Query Processing and Optimization 1151
27.7 System Architecture 1154
Bibliographical Notes 1155

Chapter 28 Oracle

28.1 Database Design and Querying Tools 1157
28.2 SQL Variations and Extensions 1158
28.3 Storage and Indexing 1162
28.4 Query Processing and Optimization 1172
28.5 Concurrency Control and Recovery 1180
28.6 System Architecture 1183
28.7 Replication, Distribution, and External Data 1188
28.8 Database Administration Tools 1189
28.9 Data Mining 1191
Bibliographical Notes 1191

Chapter 29 IBM DB2 Universal Database

29.1 Overview 1193
29.2 Database-Design Tools 1194
29.3 SQL Variations and Extensions 1195
29.4 Storage and Indexing 1200
29.5 Multidimensional Clustering 1203
29.6 Query Processing and Optimization 1207
29.7 Materialized Query Tables 1212
29.8 Autonomic Features in DB2 1214
29.9 Tools and Utilities 1215
29.10 Concurrency Control and Recovery 1217
29.11 System Architecture 1219
29.12 Replication, Distribution, and External Data 1220
29.13 Business Intelligence Features 1221
Bibliographical Notes 1222

Chapter 30 Microsoft SQL Server

30.1 Management, Design, and Querying Tools 1223
30.2 SQL Variations and Extensions 1228
30.3 Storage and Indexing 1233
30.4 Query Processing and Optimization 1236
30.5 Concurrency and Recovery 1241
30.6 System Architecture 1246
30.7 Data Access 1248
30.8 Distributed Heterogeneous Query Processing 1250
30.9 Replication 1251
30.10 Server Programming in .NET 1253
30.11 XML Support 1258
30.12 SQL Server Service Broker 1261
30.13 Business Intelligence 1263
Bibliographical Notes 1267

PART TEN ■ APPENDICES

Appendix A Detailed University Schema

A.1 Full Schema 1271
A.2 DDL 1272
A.3 Sample Data 1276

Appendix B Advanced Relational Design (contents online)

B.1 Multivalued Dependencies B1
B.3 Domain-Key Normal Form B8
B.4 Summary B10
Exercises B10
Bibliographical Notes B12

Appendix C Other Relational Query Languages (contents online)

C.1 Query-by-Example C1
C.2 Microsoft Access C9
C.3 Datalog C11
C.4 Summary C25
Exercises C26
Bibliographical Notes C30

Appendix D Network Model (contents online)

D.1 Basic Concepts D1
D.2 Data-Structure Diagrams D2
D.3 The DBTG CODASYL Model D7
D.4 DBTG Data-Retrieval Facility D13
D.5 DBTG Update Facility D20
D.6 DBTG Set-Processing Facility D22
D.7 Mapping of Networks to Files D27
D.8 Summary D31
Exercises D32
Bibliographical Notes D35

Appendix E Hierarchical Model (contents online)

E.1 Basic Concepts E1
E.2 Tree-Structure Diagrams E2
E.3 Data-Retrieval Facility E13
E.4 Update Facility E17
E.5 Virtual Records E20
E.6 Mapping of Hierarchies to Files E22
E.7 The IMS Database System E24
E.8 Summary E25
Exercises E26
Bibliographical Notes E29

Bibliography 1283

Index 1315