

GLOBAL
EDITION



Fundamentals of Database Systems

SEVENTH EDITION

Ramez Elmasri • Shamkant B. Navathe

ALWAYS LEARNING

PEARSON

Contents

Preface	7
About the Authors	30

■ part 1

Introduction to Databases ■

chapter 1 Databases and Database Users 33

1.1 Introduction	34
1.2 An Example	36
1.3 Characteristics of the Database Approach	40
1.4 Actors on the Scene	45
1.5 Workers behind the Scene	47
1.6 Advantages of Using the DBMS Approach	47
1.7 A Brief History of Database Applications	53
1.8 When Not to Use a DBMS	57
1.9 Summary	57
Review Questions	58
Exercises	58
Selected Bibliography	59

chapter 2 Database System Concepts and Architecture 61

2.1 Data Models, Schemas, and Instances	62
2.2 Three-Schema Architecture and Data Independence	66
2.3 Database Languages and Interfaces	69
2.4 The Database System Environment	72
2.5 Centralized and Client/Server Architectures for DBMSs	78
2.6 Classification of Database Management Systems	81
2.7 Summary	84
Review Questions	85
Exercises	85
Selected Bibliography	86

■ part 2

Conceptual Data Modeling and Database Design ■

chapter 3 Data Modeling Using the Entity–Relationship (ER) Model 89

- 3.1 Using High-Level Conceptual Data Models for Database Design 90
- 3.2 A Sample Database Application 92
- 3.3 Entity Types, Entity Sets, Attributes, and Keys 93
- 3.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints 102
- 3.5 Weak Entity Types 109
- 3.6 Refining the ER Design for the COMPANY Database 110
- 3.7 ER Diagrams, Naming Conventions, and Design Issues 111
- 3.8 Example of Other Notation: UML Class Diagrams 115
- 3.9 Relationship Types of Degree Higher than Two 118
- 3.10 Another Example: A UNIVERSITY Database 122
- 3.11 Summary 124
- Review Questions 126
- Exercises 126
- Laboratory Exercises 133
- Selected Bibliography 134

chapter 4 The Enhanced Entity–Relationship (EER) Model 137

- 4.1 Subclasses, Superclasses, and Inheritance 138
- 4.2 Specialization and Generalization 140
- 4.3 Constraints and Characteristics of Specialization and Generalization Hierarchies 143
- 4.4 Modeling of UNION Types Using Categories 150
- 4.5 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions 152
- 4.6 Example of Other Notation: Representing Specialization and Generalization in UML Class Diagrams 157
- 4.7 Data Abstraction, Knowledge Representation, and Ontology Concepts 158
- 4.8 Summary 165
- Review Questions 165
- Exercises 166
- Laboratory Exercises 173
- Selected Bibliography 176

■ part 3

The Relational Data Model and SQL ■

chapter 5 The Relational Data Model and Relational Database Constraints 179

- 5.1 Relational Model Concepts 180
- 5.2 Relational Model Constraints and Relational Database Schemas 187
- 5.3 Update Operations, Transactions, and Dealing with Constraint Violations 195
- 5.4 Summary 199
- Review Questions 200
- Exercises 200
- Selected Bibliography 205

chapter 6 Basic SQL 207

- 6.1 SQL Data Definition and Data Types 209
- 6.2 Specifying Constraints in SQL 214
- 6.3 Basic Retrieval Queries in SQL 217
- 6.4 INSERT, DELETE, and UPDATE Statements in SQL 228
- 6.5 Additional Features of SQL 231
- 6.6 Summary 232
- Review Questions 233
- Exercises 233
- Selected Bibliography 235

chapter 7 More SQL: Complex Queries, Triggers, Views, and Schema Modification 237

- 7.1 More Complex SQL Retrieval Queries 237
- 7.2 Specifying Constraints as Assertions and Actions as Triggers 255
- 7.3 Views (Virtual Tables) in SQL 258
- 7.4 Schema Change Statements in SQL 262
- 7.5 Summary 264
- Review Questions 266
- Exercises 266
- Selected Bibliography 268

chapter 8 The Relational Algebra and Relational Calculus 269

- 8.1 Unary Relational Operations: SELECT and PROJECT 271
- 8.2 Relational Algebra Operations from Set Theory 276

8.3 Binary Relational Operations: JOIN and DIVISION	281
8.4 Additional Relational Operations	289
8.5 Examples of Queries in Relational Algebra	295
8.6 The Tuple Relational Calculus	298
8.7 The Domain Relational Calculus	307
8.8 Summary	309
Review Questions	310
Exercises	311
Laboratory Exercises	316
Selected Bibliography	318

chapter 9 Relational Database Design by ER- and EER-to-Relational Mapping 319

9.1 Relational Database Design Using ER-to-Relational Mapping	320
9.2 Mapping EER Model Constructs to Relations	328
9.3 Summary	333
Review Questions	333
Exercises	333
Laboratory Exercises	335
Selected Bibliography	336

part 4

Database Programming Techniques ■

chapter 10 Introduction to SQL Programming Techniques 339

10.1 Overview of Database Programming Techniques and Issues	340
10.2 Embedded SQL, Dynamic SQL, and SQLJ	344
10.3 Database Programming with Function Calls and Class Libraries: SQL/CLI and JDBC	355
10.4 Database Stored Procedures and SQL/PSM	365
10.5 Comparing the Three Approaches	368
10.6 Summary	369
Review Questions	370
Exercises	370
Selected Bibliography	371

chapter 11 Web Database Programming Using PHP 373

11.1 A Simple PHP Example	374
11.2 Overview of Basic Features of PHP	376

- 11.3 Overview of PHP Database Programming 383
- 11.4 Brief Overview of Java Technologies for Database Web Programming 388
- 11.5 Summary 388
- Review Questions 399
- Exercises 389
- Selected Bibliography 389

■ part 5

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

chapter **12** Object and Object-Relational Databases 393

- 12.1 Overview of Object Database Concepts 395
- 12.2 Object Database Extensions to SQL 409
- 12.3 The ODMG Object Model and the Object Definition Language ODL 416
- 12.4 Object Database Conceptual Design 435
- 12.5 The Object Query Language OQL 438
- 12.6 Overview of the C++ Language Binding in the ODMG Standard 447
- 12.7 Summary 448
- Review Questions 450
- Exercises 451
- Selected Bibliography 452

chapter **13** XML: Extensible Markup Language 455

- 13.1 Structured, Semistructured, and Unstructured Data 456
- 13.2 XML Hierarchical (Tree) Data Model 460
- 13.3 XML Documents, DTD, and XML Schema 463
- 13.4 Storing and Extracting XML Documents from Databases 472
- 13.5 XML Languages 473
- 13.6 Extracting XML Documents from Relational Databases 477
- 13.7 XML/SQL: SQL Functions for Creating XML Data 483
- 13.8 Summary 485
- Review Questions 486
- Exercises 486
- Selected Bibliography 486

■ part 6

Database Design Theory and Normalization ■

chapter 14 Basics of Functional Dependencies and Normalization for Relational Databases 489

- 14.1 Informal Design Guidelines for Relation Schemas 491
- 14.2 Functional Dependencies 501
- 14.3 Normal Forms Based on Primary Keys 504
- 14.4 General Definitions of Second and Third Normal Forms 513
- 14.5 Boyce-Codd Normal Form 517
- 14.6 Multivalued Dependency and Fourth Normal Form 521
- 14.7 Join Dependencies and Fifth Normal Form 524
- 14.8 Summary 525
- Review Questions 526
- Exercises 527
- Laboratory Exercises 531
- Selected Bibliography 532

chapter 15 Relational Database Design Algorithms and Further Dependencies 533

- 15.1 Further Topics in Functional Dependencies: Inference Rules, Equivalence, and Minimal Cover 535
- 15.2 Properties of Relational Decompositions 543
- 15.3 Algorithms for Relational Database Schema Design 549
- 15.4 About Nulls, Dangling Tuples, and Alternative Relational Designs 553
- 15.5 Further Discussion of Multivalued Dependencies and 4NF 557
- 15.6 Other Dependencies and Normal Forms 560
- 15.7 Summary 563
- Review Questions 564
- Exercises 565
- Laboratory Exercises 566
- Selected Bibliography 567

part 7

File Structures, Hashing, Indexing, and Physical Database Design

chapter 16 Disk Storage, Basic File Structures, Hashing, and Modern Storage Architectures 571

- 16.1 Introduction 572
- 16.2 Secondary Storage Devices 577
- 16.3 Buffering of Blocks 586
- 16.4 Placing File Records on Disk 590
- 16.5 Operations on Files 594
- 16.6 Files of Unordered Records (Heap Files) 597
- 16.7 Files of Ordered Records (Sorted Files) 598
- 16.8 Hashing Techniques 602
- 16.9 Other Primary File Organizations 612
- 16.10 Parallelizing Disk Access Using RAID Technology 614
- 16.11 Modern Storage Architectures 618
- 16.12 Summary 622
- Review Questions 623
- Exercises 625
- Selected Bibliography 628

chapter 17 Indexing Structures for Files and Physical Database Design 631

- 17.1 Types of Single-Level Ordered Indexes 632
- 17.2 Multilevel Indexes 643
- 17.3 Dynamic Multilevel Indexes Using B-Trees and B⁺-Trees 647
- 17.4 Indexes on Multiple Keys 661
- 17.5 Other Types of Indexes 663
- 17.6 Some General Issues Concerning Indexing 668
- 17.7 Physical Database Design in Relational Databases 673
- 17.8 Summary 676
- Review Questions 677
- Exercises 678
- Selected Bibliography 680

part 8

Query Processing and Optimization

chapter 18 Strategies for Query Processing 685

- 18.1 Translating SQL Queries into Relational Algebra and Other Operators 687
- 18.2 Algorithms for External Sorting 690
- 18.3 Algorithms for SELECT Operation 693
- 18.4 Implementing the JOIN Operation 698
- 18.5 Algorithms for PROJECT and Set Operations 706
- 18.6 Implementing Aggregate Operations and Different Types of JOINS 708
- 18.7 Combining Operations Using Pipelining 711
- 18.8 Parallel Algorithms for Query Processing 713
- 18.9 Summary 718
- Review Questions 718
- Exercises 719
- Selected Bibliography 719

chapter 19 Query Optimization 721

- 19.1 Query Trees and Heuristics for Query Optimization 722
- 19.2 Choice of Query Execution Plans 731
- 19.3 Use of Selectivities in Cost-Based Optimization 740
- 19.4 Cost Functions for SELECT Operation 744
- 19.5 Cost Functions for the JOIN Operation 747
- 19.6 Example to Illustrate Cost-Based Query Optimization 756
- 19.7 Additional Issues Related to Query Optimization 758
- 19.8 An Example of Query Optimization in Data Warehouses 761
- 19.9 Overview of Query Optimization in Oracle 763
- 19.10 Semantic Query Optimization 767
- 19.11 Summary 768
- Review Questions 769
- Exercises 770
- Selected Bibliography 770

part 9

Transaction Processing, Concurrency Control, and Recovery

chapter 20 Introduction to Transaction Processing Concepts and Theory 775

- 20.1 Introduction to Transaction Processing 776
- 20.2 Transaction and System Concepts 783
- 20.3 Desirable Properties of Transactions 787
- 20.4 Characterizing Schedules Based on Recoverability 789
- 20.5 Characterizing Schedules Based on Serializability 793
- 20.6 Transaction Support in SQL 803
- 20.7 Summary 806
- Review Questions 807
- Exercises 807
- Selected Bibliography 809

chapter 21 Concurrency Control Techniques 811

- 21.1 Two-Phase Locking Techniques for Concurrency Control 812
- 21.2 Concurrency Control Based on Timestamp Ordering 822
- 21.3 Multiversion Concurrency Control Techniques 825
- 21.4 Validation (Optimistic) Techniques and Snapshot Isolation Concurrency Control 828
- 21.5 Granularity of Data Items and Multiple Granularity Locking 830
- 21.6 Using Locks for Concurrency Control in Indexes 835
- 21.7 Other Concurrency Control Issues 836
- 21.8 Summary 837
- Review Questions 838
- Exercises 839
- Selected Bibliography 840

chapter 22 Database Recovery Techniques 843

- 22.1 Recovery Concepts 844
- 22.2 NO-UNDO/REDO Recovery Based on Deferred Update 851
- 22.3 Recovery Techniques Based on Immediate Update 853

22.4 Shadow Paging	856
22.5 The ARIES Recovery Algorithm	857
22.6 Recovery in Multidatabase Systems	861
22.7 Database Backup and Recovery from Catastrophic Failures	862
22.8 Summary	863
Review Questions	864
Exercises	865
Selected Bibliography	868

■ part 10

Distributed Databases, NOSQL Systems, and Big Data ■

chapter 23 Distributed Database Concepts	871
23.1 Distributed Database Concepts	872
23.2 Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design	877
23.3 Overview of Concurrency Control and Recovery in Distributed Databases	884
23.4 Overview of Transaction Management in Distributed Databases	887
23.5 Query Processing and Optimization in Distributed Databases	889
23.6 Types of Distributed Database Systems	895
23.7 Distributed Database Architectures	898
23.8 Distributed Catalog Management	905
23.9 Summary	906
Review Questions	907
Exercises	908
Selected Bibliography	910
chapter 24 NOSQL Databases and Big Data Storage Systems	913
24.1 Introduction to NOSQL Systems	914
24.2 The CAP Theorem	918
24.3 Document-Based NOSQL Systems and MongoDB	920
24.4 NOSQL Key-Value Stores	925
24.5 Column-Based or Wide Column NOSQL Systems	930
24.6 NOSQL Graph Databases and Neo4j	933
24.7 Summary	939
Review Questions	939
Selected Bibliography	940

chapter **25** Big Data Technologies Based on MapReduce and Hadoop 941

- 25.1 What Is Big Data? 944
- 25.2 Introduction to MapReduce and Hadoop 946
- 25.3 Hadoop Distributed File System (HDFS) 951
- 25.4 MapReduce: Additional Details 956
- 25.5 Hadoop v2 alias YARN 966
- 25.6 General Discussion 974
- 25.7 Summary 983
- Review Questions 984
- Selected Bibliography 986

■ part 11

Advanced Database Models, Systems, and Applications ■

chapter **26** Enhanced Data Models: Introduction to Active, Temporal, Spatial, Multimedia, and Deductive Databases 991

- 26.1 Active Database Concepts and Triggers 993
- 26.2 Temporal Database Concepts 1004
- 26.3 Spatial Database Concepts 1017
- 26.4 Multimedia Database Concepts 1024
- 26.5 Introduction to Deductive Databases 1029
- 26.6 Summary 1042
- Review Questions 1044
- Exercises 1045
- Selected Bibliography 1048

chapter **27** Introduction to Information Retrieval and Web Search 1051

- 27.1 Information Retrieval (IR) Concepts 1052
- 27.2 Retrieval Models 1059
- 27.3 Types of Queries in IR Systems 1065
- 27.4 Text Preprocessing 1067
- 27.5 Inverted Indexing 1070
- 27.6 Evaluation Measures of Search Relevance 1074
- 27.7 Web Search and Analysis 1077

27.8 Trends in Information Retrieval	1087
27.9 Summary	1093
Review Questions	1094
Selected Bibliography	1096
chapter 28 Data Mining Concepts	1099
28.1 Overview of Data Mining Technology	1100
28.2 Association Rules	1103
28.3 Classification	1115
28.4 Clustering	1118
28.5 Approaches to Other Data Mining Problems	1121
28.6 Applications of Data Mining	1124
28.7 Commercial Data Mining Tools	1124
28.8 Summary	1127
Review Questions	1127
Exercises	1128
Selected Bibliography	1129
chapter 29 Overview of Data Warehousing and OLAP	1131
29.1 Introduction, Definitions, and Terminology	1132
29.2 Characteristics of Data Warehouses	1133
29.3 Data Modeling for Data Warehouses	1135
29.4 Building a Data Warehouse	1141
29.5 Typical Functionality of a Data Warehouse	1144
29.6 Data Warehouse versus Views	1145
29.7 Difficulties of Implementing Data Warehouses	1146
29.8 Summary	1147
Review Questions	1147
Selected Bibliography	1148

■ part 12

Additional Database Topics: Security ■

chapter 30 Database Security	1151
30.1 Introduction to Database Security Issues	1152
30.2 Discretionary Access Control Based on Granting and Revoking Privileges	1159
30.3 Mandatory Access Control and Role-Based Access Control for Multilevel Security	1164

30.4 SQL Injection	1173
30.5 Introduction to Statistical Database Security	1176
30.6 Introduction to Flow Control	1177
30.7 Encryption and Public Key Infrastructures	1179
30.8 Privacy Issues and Preservation	1183
30.9 Challenges to Maintaining Database Security	1184
30.10 Oracle Label-Based Security	1185
30.11 Summary	1188
Review Questions	1189
Exercises	1190
Selected Bibliography	1191

appendix **A** **Alternative Diagrammatic Notations for ER Models** 1193

appendix **B** **Parameters of Disks** 1197

appendix **C** **Overview of the QBE Language** 1201

C.1 Basic Retrievals in QBE 1201

C.2 Grouping, Aggregation, and Database Modification in QBE 1205

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

(located on the Companion Website at
www.pearsonglobaleditions.com/Elmasri)

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

(located on the Companion Website at
www.pearsonglobaleditions.com/Elmasri)

Bibliography 1209

Index 1245