

Database Systems:

Models, Languages, Design,
and Application Programming

Ramez Elmasri

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

Shamkant B. Navathe

*College of Computing
Georgia Institute of Technology*

PEARSON

Boston Columbus Indianapolis New York San Francisco 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

Contents

chapter 1	Introduction to Databases	1
1.1	Introduction	2
1.2	An Example	4
1.3	Characteristics of the Database Approach	7
1.4	Actors on the Scene	12
1.5	Workers behind the Scene	14
1.6	Advantages of Using the DBMS Approach	15
1.7	A Brief History of Database Applications	21
1.8	When Not to Use a DBMS	24
1.9	Summary	25
	Review Questions	25
	Exercises	26
	Selected Bibliography	26
chapter 2	Overview of Database Languages and Architectures	27
2.1	Data Models, Schemas, and Instances	28
2.2	Three-Schema Architecture and Data Independence	31
2.3	Database Languages and Interfaces	34
2.4	The Database System Environment	38
2.5	Centralized and Client/Server Architectures for DBMSs	42
2.6	Classification of Database Management Systems	47
2.7	Summary	50
	Review Questions	51
	Exercises	52
	Selected Bibliography	53
chapter 3	The Basic (Flat) Relational Model	55
3.1	Relational Model Concepts	56
3.2	Relational Model Constraints and Relational Database Schemas	63
3.3	Update Operations, Transactions, and Dealing with Constraint Violations	71
3.4	Summary	75

Review Questions 76
Exercises 76
Selected Bibliography 81

chapter **4** **SQL: Data Definition, Constraints, and Basic Queries and Updates 83**

4.1 SQL Data Definition and Data Types 85
4.2 Specifying Constraints in SQL 90
4.3 Basic Retrieval Queries in SQL 93
4.4 INSERT, DELETE, and UPDATE Statements in SQL 103
4.5 Additional Features of SQL 106
4.6 Summary 107
Review Questions 108
Exercises 108
Selected Bibliography 110

chapter **5** **SQL: Advanced Queries, Assertions, Triggers, and Views 111**

5.1 More Complex SQL Retrieval Queries 111
5.2 Specifying Constraints as Assertions and Actions as Triggers 127
5.3 Views (Virtual Tables) in SQL 129
5.4 Schema Change Statements in SQL 133
5.5 Summary 135
Review Questions 137
Exercises 137
Selected Bibliography 139

chapter **6** **Formal Relational Languages: The Algebra and Calculus 141**

6.1 Unary Relational Operations: SELECT and PROJECT 143
6.2 Relational Algebra Operations from Set Theory 148
6.3 Binary Relational Operations: JOIN and DIVISION 153
6.4 Additional Relational Operations 161
6.5 Examples of Queries in Relational Algebra 167
6.6 The Tuple Relational Calculus 170
6.7 The Domain Relational Calculus 179
6.8 Summary 181

Review Questions	182
Exercises	183
Laboratory Exercises	188
Selected Bibliography	190

chapter **7** **Conceptual Data Modeling Using Entities and Relationships** 193

7.1 Using High-Level Conceptual Data Models for Database Design	195
7.2 A Sample Database Application	197
7.3 Entity Types, Entity Sets, Attributes, and Keys	198
7.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints	206
7.5 Weak Entity Types	213
7.6 Refining the ER Design for the COMPANY Database	214
7.7 ER Diagrams, Naming Conventions, and Design Issues	215
7.8 Relationship Types of Degree Higher than Two	219
7.9 Subclasses, Superclasses, and Inheritance	224
7.10 Specialization and Generalization in EER	226
7.11 Constraints and Characteristics of Specialization and Generalization Hierarchies	229
7.12 Modeling of UNION Types Using Categories in EER	236
7.13 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions	238
7.14 Data Abstraction, Knowledge Representation, and Ontology Concepts	243
7.15 Summary	249
Review Questions	250
Exercises	252
Laboratory Exercises	263
Selected Bibliography	267

chapter **8** **Mapping a Conceptual Design into a Logical Design** 269

8.1 Relational Database Design Using ER-to-Relational Mapping	270
8.2 Mapping EER Model Constructs to Relations	278
8.3 Summary	283
Review Questions	283
Exercises	283
Laboratory Exercises	285
Selected Bibliography	286

chapter **9** UML for Database Application Design 287

- 9.1 Basics of UML Class Diagrams 289
- 9.2 Representing Specialization and Generalization in UML Class Diagrams 291
- 9.3 Use of UML Diagrams as an Aid to Database Design Specification 293
- 9.4 Rational Rose: A UML-Based Design Tool 301
- 9.5 The Role of Information Systems in Organizations 306
- 9.6 The Database Design and Implementation Process 311
- 9.7 Automated Database Design Tools 330
- 9.8 Summary 334
- Review Questions 334
- Exercises 336
- Selected Bibliography 338

chapter **10** Object and Object-Relational Databases: Concepts, Models, Languages, and Standards 341

- 10.1 Overview of Object Database Concepts 343
- 10.2 Object-Relational Features: Object Database Extensions to SQL 357
- 10.3 The ODMG Object Model and the Object Definition Language ODL 364
- 10.4 Object Database Conceptual Design 383
- 10.5 The Object Query Language OQL 386
- 10.6 Overview of the C++ Language Binding in the ODMG Standard 395
- 10.7 Summary 396
- Review Questions 397
- Exercises 399
- Selected Bibliography 400

chapter **11** XML: Concepts, Languages, and Standards 403

- 11.1 Structured, Semistructured, and Unstructured Data 404
- 11.2 XML Hierarchical (Tree) Data Model 408
- 11.3 XML Documents, DTD, and XML Schema 411
- 11.4 Storing and Extracting XML Documents from Databases 419
- 11.5 XML Languages 420
- 11.6 Extracting XML Documents from Relational Databases 424

11.7 Summary	430
Review Questions	430
Exercises	431
Selected Bibliography	431

chapter **12** SQL Application Programming Using C and Java 433

12.1 Database Programming: Techniques and Issues	434
12.2 Embedded SQL, Dynamic SQL, and SQLJ	437
12.3 Database Programming with Function Calls: SQL/CLI and JDBC	450
12.4 Database Stored Procedures and SQL/PSM	459
12.5 Comparing the Three Approaches	462
12.6 Summary	463
Review Questions	464
Exercises	464
Selected Bibliography	465

chapter **13** SQL Web Programming Using C PHP 467

13.1 A Simple PHP Example	468
13.2 Overview of Basic Features of PHP	470
13.3 Overview of PHP Database Programming	477
13.4 Summary	482
Review Questions	482
Exercises	483
Selected Bibliography	483

chapter **14** Database Design Theory: Introduction to Normalization Using Functional and Multivalued Dependencies 485

14.1 Informal Design Guidelines for Relation Schemas	487
14.2 Functional Dependencies	497
14.3 Normal Forms Based on Primary Keys	500
14.4 General Definitions of Second and Third Normal Forms	509
14.5 Boyce-Codd Normal Form	513
14.6 Multivalued Dependency and Fourth Normal Form	515
14.7 Join Dependencies and Fifth Normal Form	518
14.8 Summary	519

Review Questions	520
Exercises	521
Laboratory Exercise	526
Selected Bibliography	526

chapter **15** Database Design Theory: Normalization Algorithms 527

15.1 Further Topics in Functional Dependencies: Inference Rules, Equivalence, and Minimal Cover	529
15.2 Properties of Relational Decompositions	535
15.3 Algorithms for Relational Database Schema Design	541
15.4 About Nulls, Dangling Tuples, and Alternative Relational Designs	547
15.5 Further Discussion of Multivalued Dependencies and 4NF	551
15.6 Other Dependencies and Normal Forms	555
15.7 Summary	559
Review Questions	560
Exercises	561
Laboratory Exercises	562
Selected Bibliography	563

chapter **16** Database File Organizations: Unordered, Ordered, and Hashed Files of Records 565

16.1 Introduction	566
16.2 Secondary Storage Devices	569
16.3 Buffering of Blocks	575
16.4 Placing File Records on Disk	576
16.5 Operations on Files	581
16.6 Files of Unordered Records (Heap Files)	583
16.7 Files of Ordered Records (Sorted Files)	585
16.8 Hashing Techniques	588
16.9 Other Primary File Organizations	598
16.10 Parallelizing Disk Access Using RAID Technology	599
16.11 New Storage Systems	603
16.12 Summary	606
Review Questions	607
Exercises	608
Selected Bibliography	612

chapter **17** Database File Indexing Techniques, B-Trees, and B+-Trees **613**

- 17.1 Types of Single-Level Ordered Indexes 614
- 17.2 Multilevel Indexes 625
- 17.3 Dynamic Multilevel Indexes Using B-Trees and B+-Trees 628
- 17.4 Indexes on Multiple Keys 642
- 17.5 Other Types of Indexes 645
- 17.6 Some General Issues Concerning Indexing 650
- 17.7 Summary 652
- Review Questions 653
- Exercises 654
- Selected Bibliography 656

chapter **18** Introduction to Query Processing and Query Optimization Techniques **659**

- 18.1 Translating SQL Queries into Relational Algebra 661
- 18.2 Algorithms for External Sorting 662
- 18.3 Algorithms for SELECT and JOIN Operations 665
- 18.4 Algorithms for PROJECT and Set Operations 676
- 18.5 Implementing Aggregate Operations and OUTER JOINS 678
- 18.6 Combining Operations Using Pipelining 680
- 18.7 Using Heuristics in Query Optimization 680
- 18.8 Using Selectivity and Cost Estimates in Query Optimization 690
- 18.9 Overview of Query Optimization in Oracle 701
- 18.10 Semantic Query Optimization 702
- 18.11 Summary 703
- Review Questions 703
- Exercises 704
- Selected Bibliography 705

chapter **19** Introduction to Database Tuning and Physical Design Issues **707**

- 19.1 Physical Database Design in Relational Databases 707
- 19.2 An Overview of Database Tuning in Relational Systems 713
- 19.3 Summary 719
- Review Questions 719
- Selected Bibliography 720

chapter **20** Foundations of Database Transaction Processing 721

- 20.1 Introduction to Transaction Processing 722
- 20.2 Transaction and System Concepts 729
- 20.3 Desirable Properties of Transactions 732
- 20.4 Characterizing Schedules Based on Recoverability 733
- 20.5 Characterizing Schedules Based on Serializability 737
- 20.6 Transaction Support in SQL 748
- 20.7 Summary 750
- Review Questions 750
- Exercises 751
- Selected Bibliography 753

chapter **21** Introduction to Protocols for Concurrency Control in Databases 755

- 21.1 Two-Phase Locking Techniques for Concurrency Control 756
- 21.2 Concurrency Control Based on Timestamp Ordering 766
- 21.3 Multiversion Concurrency Control Techniques 769
- 21.4 Validation (Optimistic) Concurrency Control Techniques 772
- 21.5 Granularity of Data Items and Multiple Granularity Locking 773
- 21.6 Using Locks for Concurrency Control in Indexes 776
- 21.7 Other Concurrency Control Issues 778
- 21.8 Summary 780
- Review Questions 781
- Exercises 782
- Selected Bibliography 782

chapter **22** Introduction to Database Recovery Protocols 785

- 22.1 Recovery Concepts 786
- 22.2 NO-UNDO/REDO Recovery Based on Deferred Update 793
- 22.3 Recovery Techniques Based on Immediate Update 795
- 22.4 Shadow Paging 798
- 22.5 The ARIES Recovery Algorithm 799
- 22.6 Recovery in Multidatabase Systems 803
- 22.7 Database Backup and Recovery from Catastrophic Failures 804
- 22.8 Summary 805
- Review Questions 806
- Exercises 807
- Selected Bibliography 810

chapter 23	Emerging Database Technologies and Applications	811
23.1	Mobile Data Management	812
23.2	Multimedia Data Management	823
23.3	Geographic Information Systems (GIS)	834
23.4	Biological and Genomic Databases and Emerging Applications	848
	Selected Bibliography	861

chapter 24	Advanced Database Models and Applications	867
-------------------	--	------------

24.1	Active Database Concepts and Triggers	869
24.2	Temporal Database Concepts	879
24.3	Spatial Database Concepts	893
24.4	Multimedia Database Concepts	901
24.5	Introduction to Deductive Databases	906
24.6	Summary	919
	Review Questions	921
	Exercises	922
	Selected Bibliography	925

chapter 25	Introduction to Database Security	929
-------------------	--	------------

25.1	Introduction to Database Security Issues	930
25.2	Discretionary Access Control Based on Granting and Revoking Privileges	936
25.3	Mandatory Access Control and Role-Based Access Control for Multilevel Security	941
25.4	SQL Injection	949
25.5	Introduction to Statistical Database Security	953
25.6	Introduction to Flow Control	954
25.7	Encryption and Public Key Infrastructures	956
25.8	Privacy Issues and Preservation	960
25.9	Challenges of Database Security	961
25.10	Oracle Label-Based Security	962
25.11	Summary	964
	Review Questions	966
	Exercises	967
	Selected Bibliography	968

chapter 26	Introduction to Distributed Databases	971
26.1	Distributed Database Concepts	972
26.2	Types of Distributed Database Systems	977
26.3	Distributed Database Architectures	981
26.4	Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design	988
26.5	Query Processing and Optimization in Distributed Databases	995
26.6	Overview of Transaction Management in Distributed Databases	1001
26.7	Overview of Concurrency Control and Recovery in Distributed Databases	1003
26.8	Distributed Catalog Management	1007
26.9	Current Trends in Distributed Databases	1008
26.10	Distributed Databases in Oracle	1009
26.11	Summary	1013
	Review Questions	1015
	Exercises	1016
	Selected Bibliography	1018
chapter 27	Introduction to Information Retrieval	1023
27.1	Information Retrieval (IR) Concepts	1024
27.2	Retrieval Models	1031
27.3	Types of Queries in IR Systems	1037
27.4	Text Preprocessing	1039
27.5	Inverted Indexing	1042
27.6	Evaluation Measures of Search Relevance	1044
27.7	Web Search and Analysis	1048
27.8	Trends in Information Retrieval	1058
27.9	Summary	1060
	Review Questions	1061
	Selected Bibliography	1063
appendix A	Alternative Diagrammatic Notations for ER Models	1065
appendix B	Parameters of Disks	1069

appendix **C** **Overview of the QBE Language** 1073

- C.1 Basic Retrievals in QBE 1073
- C.2 Grouping, Aggregation, and Database
Modification in QBE 1077

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

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

Selected Bibliography 1081

Index 1115