

# Database Systems:

Models, Languages, Design,  
and Application Programming

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*

PEARSON

Chennai • Delhi

# 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 this 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 conceptual modeling using entities and relationships (ER and EER) has been consolidated into a single chapter (Chapter 7), and the discussion on UML is consolidated in Chapter 9
- 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 10)
- The presentation of XML has been expanded and updated, and moved earlier in the book to Chapter 11
- The chapters on normalization theory have been reorganized so that the first chapter (Chapter 14) focuses on intuitive normalization concepts, while the second chapter (Chapter 15) focuses on the formal theories and normalization algorithms.

# Brief Contents

- chapter **1** Introduction to Databases 1
- chapter **2** Overview of Database Languages and Architectures 29
- chapter **3** The Basic (Flat) Relational Model 57
- chapter **4** SQL: Data Definition, Constraints, and Basic Queries and Updates 85
- chapter **5** SQL: Advanced Queries, Assertions, Triggers, and Views 115
- chapter **6** Formal Relational Languages: The Algebra and Calculus 145
- chapter **7** Conceptual Data Modeling Using Entities and Relationships 197
- chapter **8** Mapping a Conceptual Design into a Logical Design 271
- chapter **9** UML for Database Application Design 289
- chapter **10** Object and Object-Relational Databases: Concepts, Models, Languages, and Standards 343
- chapter **11** XML: Concepts, Languages, and Standards 405
- chapter **12** SQL Application Programming Using C and Java 435
- chapter **13** SQL Web Programming Using C PHP 469
- chapter **14** Database Design Theory: Introduction to Normalization Using Functional and Multivalued Dependencies 487
- chapter **15** Database Design Theory: Normalization Algorithms 531
- chapter **16** Database File Organizations: Unordered, Ordered, and Hashed Files of Records 569
- chapter **17** Database File Indexing Techniques, B-Trees, and B+-Trees 617
- chapter **18** Introduction to Query Processing and Query Optimization Techniques 663

chapter <b>19</b>	Introduction to Database Tuning and Physical Design Issues	711
chapter <b>20</b>	Foundations of Database Transaction Processing	725
chapter <b>21</b>	Introduction to Protocols for Concurrency Control in Databases	759
chapter <b>22</b>	Introduction to Database Recovery Protocols	789
chapter <b>23</b>	Emerging Database Technologies and Applications	817
chapter <b>24</b>	Advanced Database Models and Applications	873
chapter <b>25</b>	Introduction to Database Security	935
chapter <b>26</b>	Introduction to Distributed Databases	977
chapter <b>27</b>	Introduction to Information Retrieval	1029
chapter <b>28</b>	Data Mining Concepts	1071
chapter <b>29</b>	Overview of Data Warehousing and OLAP	1103
appendix <b>A</b>	Alternative Diagrammatic Notations for ER Models	1121
appendix <b>B</b>	Parameters of Disks	1125
appendix <b>C</b>	Overview of the QBE Language	1129
appendix <b>D</b>	Overview of the Hierarchical Data Model (located on the Companion Web site at <a href="http://www.pearsoned.co.in/ramezelmasri">http://www.pearsoned.co.in/ramezelmasri</a> )	
appendix <b>E</b>	Overview of the Network Data Model (located on the Companion Web site at <a href="http://www.pearsoned.co.in/ramezelmasri">http://www.pearsoned.co.in/ramezelmasri</a> )	
	Selected Bibliography	1137
	Index	1173

# Contents

<b>chapter 1</b>	<b>Introduction to Databases</b>	<b>1</b>
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	27
<b>chapter 2</b>	<b>Overview of Database Languages and Architectures</b>	<b>29</b>
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
<b>chapter 3</b>	<b>The Basic (Flat) Relational Model</b>	<b>57</b>
3.1	Relational Model Concepts	58
3.2	Relational Model Constraints and Relational Database Schemas	65
3.3	Update Operations, Transactions, and Dealing with Constraint Violations	73

3.4 Summary	77
Review Questions	78
Exercises	79
Selected Bibliography	83

**chapter 4 SQL: Data Definition, Constraints, and Basic Queries and Updates 85**

4.1 SQL Data Definition and Data Types	87
4.2 Specifying Constraints in SQL	92
4.3 Basic Retrieval Queries in SQL	95
4.4 INSERT, DELETE, and UPDATE Statements in SQL	105
4.5 Additional Features of	108
4.6 Summary	109
Review Questions	110
Exercises	110
Selected Bibliography	113

**chapter 5 SQL: Advanced Queries, Assertions, Triggers, and Views 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 Formal Relational Languages: The Algebra and 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	194
Selected Bibliography	196

## chapter **7** Conceptual Data Modeling Using Entities and Relationships 197

7.1 Using High-Level Conceptual Data Models for Database Design	199
7.2 A Sample Database Application	201
7.3 Entity Types, Entity Sets, Attributes, and Keys	202
7.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints	210
7.5 Weak Entity Types	217
7.6 Refining the ER Design for the COMPANY Database	218
7.7 ER Diagrams, Naming Conventions, and Design Issues	219
7.8 Relationship Types of Degree Higher than Two	223
7.9 Subclasses, Superclasses, and Inheritance	228
7.10 Specialization and Generalization in EER	230
7.11 Constraints and Characteristics of Specialization and Generalization Hierarchies	233
7.12 Modeling of UNION Types Using Categories in EER	240
7.13 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions	242
7.14 Data Abstraction, Knowledge Representation, and Ontology Concepts	247
7.15 Summary	253
Review Questions	254
Exercises	256
Laboratory Exercises	265
Selected Bibliography	269

## chapter **8** Mapping a Conceptual Design into a Logical Design 271

8.1 Relational Database Design Using ER-to-Relational Mapping	272
8.2 Mapping EER Model Constructs to Relations	280
8.3 Summary	285
Review Questions	285
Exercises	285
Laboratory Exercises	287
Selected Bibliography	288

**chapter 9 UML for Database Application Design 289**

- 9.1 Basics of UML Class Diagrams 291
- 9.2 Representing Specialization and Generalization in UML Class Diagrams 293
- 9.3 Use of UML Diagrams as an Aid to Database Design Specification 295
- 9.4 Rational Rose: A UML-Based Design Tool 303
- 9.5 The Role of Information Systems in Organizations 308
- 9.6 The Database Design and Implementation Process 313
- 9.7 Automated Database Design Tools 332
- 9.8 Summary 336
- Review Questions 336
- Exercises 338
- Selected Bibliography 341

**chapter 10 Object and Object-Relational Databases: Concepts, Models, Languages, and Standards 343**

- 10.1 Overview of Object Database Concepts 345
- 10.2 Object-Relational Features: Object Database Extensions to SQL 359
- 10.3 The ODMG Object Model and the Object Definition Language ODL 366
- 10.4 Object Database Conceptual Design 385
- 10.5 The Object Query Language OQL 388
- 10.6 Overview of the C++ Language Binding in the ODMG Standard 397
- 10.7 Summary 398
- Review Questions 399
- Exercises 401
- Selected Bibliography 402

**chapter 11 XML: Concepts, Languages, and Standards 405**

- 11.1 Structured, Semistructured, and Unstructured Data 406
- 11.2 XML Hierarchical (Tree) Data Model 410
- 11.3 XML Documents, DTD, and XML Schema 413
- 11.4 Storing and Extracting XML Documents from Databases 421
- 11.5 XML Languages 422
- 11.6 Extracting XML Documents from Relational Databases 426
- 11.7 Summary 432



Review Questions	432
Exercises	433
Selected Bibliography	434

## chapter **12** SQL Application Programming Using C and Java     **435**

12.1 Database Programming: Techniques and Issues	436
12.2 Embedded SQL, Dynamic SQL, and SQLJ	439
12.3 Database Programming with Function Calls: SQL/CLI and JDBC	452
12.4 Database Stored Procedures and SQL/PSM	461
12.5 Comparing the Three Approaches	464
12.6 Summary	465
Review Questions	466
Exercises	466
Selected Bibliography	467

## chapter **13** SQL Web Programming Using C PHP     **469**

13.1 A Simple PHP Example	470
13.2 Overview of Basic Features of PHP	472
13.3 Overview of PHP Database Programming	479
13.4 Summary	484
Review Questions	484
Exercises	485
Selected Bibliography	485

## chapter **14** Database Design Theory: Introduction to Normalization Using Functional and Multivalued Dependencies     **487**

14.1 Informal Design Guidelines for Relation Schema	489
14.2 Functional Dependencies	499
14.3 Normal Forms Based on Primary Keys	502
14.4 General Definitions of Second and Third Normal Forms	511
14.5 Boyce-Codd Normal Form	515
14.6 Multivalued Dependency and Fourth Normal Form	517
14.7 Join Dependencies and Fifth Normal Form	520
14.8 Summary	521
Review Questions	522
Exercises	524

Laboratory Exercise 528  
Selected Bibliography 529

chapter **15** Database Design Theory: Normalization Algorithms 531

15.1 Further Topics in Functional Dependencies: Inference Rules, Equivalence, and Minimal Cover 533  
15.2 Properties of Relational Decompositions 539  
15.3 Algorithms for Relational Database Schema Design 545  
15.4 About Nulls, Dangling Tuples, and Alternative Relational Designs 551  
15.5 Further Discussion of Multivalued Dependencies and 4NF 555  
15.6 Other Dependencies and Normal Forms 559  
15.7 Summary 563  
Review Questions 564  
Exercises 565  
Laboratory Exercises 566  
Selected Bibliography 567

chapter **16** Database File Organizations: Unordered, Ordered, and Hashed Files of Records 569

16.1 Introduction 570  
16.2 Secondary Storage Devices 573  
16.3 Buffering of Blocks 579  
16.4 Placing File Records on Disk 580  
16.5 Operations on Files 585  
16.6 Files of Unordered Records (Heap Files) 587  
16.7 Files of Ordered Records (Sorted Files) 589  
16.8 Hashing Techniques 592  
16.9 Other Primary File Organizations 602  
16.10 Parallelizing Disk Access Using RAID Technology 603  
16.11 New Storage Systems 607  
16.12 Summary 610  
Review Questions 611  
Exercises 612  
Selected Bibliography 616

chapter **17** Database File Indexing Techniques, B-Trees, and B<sup>+</sup>-Trees 617

17.1 Types of Single-Level Ordered Indexes 618

17.2	Multilevel Indexes	629
17.3	Dynamic Multilevel Indexes Using B-Trees and B+-Trees	632
17.4	Indexes on Multiple Keys	646
17.5	Other Types of Indexes	649
17.6	Some General Issues Concerning Indexing	654
17.7	Summary	656
	Review Questions	657
	Exercises	658
	Selected Bibliography	660

## chapter **18** Introduction to Query Processing and Query Optimization Techniques    **663**

18.1	Translating SQL Queries into Relational Algebra	665
18.2	Algorithms for External Sorting	666
18.3	Algorithms for SELECT and JOIN Operations	669
18.4	Algorithms for PROJECT and Set Operations	680
18.5	Implementing Aggregate Operations and OUTER JOINS	682
18.6	Combining Operations Using Pipelining	684
18.7	Using Heuristics in Query Optimization	684
18.8	Using Selectivity and Cost Estimates in Query Optimization	694
18.9	Overview of Query Optimization in Oracle	705
18.10	Semantic Query Optimization	706
18.11	Summary	707
	Review Questions	707
	Exercises	708
	Selected Bibliography	709

## chapter **19** Introduction to Database Tuning and Physical Design Issues    **711**

19.1	Physical Database Design in Relational Databases	711
19.2	An Overview of Database Tuning in Relational Systems	717
19.3	Summary	723
	Review Questions	723
	Selected Bibliography	724

## chapter **20** Foundations of Database Transaction Processing    **725**

20.1	Introduction to Transaction Processing	726
20.2	Transaction and System Concepts	733

20.3	Desirable Properties of Transactions	736
20.4	Characterizing Schedules Based on Recoverability	737
20.5	Characterizing Schedules Based on Serializability	741
20.6	Transaction Support in SQL	752
20.7	Summary	754
	Review Questions	754
	Exercises	756
	Selected Bibliography	757

chapter **21** Introduction to Protocols for Concurrency Control in Databases 759

21.1	Two-Phase Locking Techniques for Concurrency Control	760
21.2	Concurrency Control Based on Timestamp Ordering	770
21.3	Multiversion Concurrency Control Techniques	773
21.4	Validation (Optimistic) Concurrency Control Techniques	776
21.5	Granularity of Data Items and Multiple Granularity Locking	777
21.6	Using Locks for Concurrency Control in Indexes	780
21.7	Other Concurrency Control Issues	782
21.8	Summary	784
	Review Questions	785
	Exercises	786
	Selected Bibliography	787

chapter **22** Introduction to Database Recovery Protocols 789

22.1	Recovery Concepts	790
22.2	NO-UNDO/REDO Recovery Based on Deferred Update	797
22.3	Recovery Techniques Based on Immediate Update	799
22.4	Shadow Paging	802
22.5	The ARIES Recovery Algorithm	803
22.6	Recovery in Multidatabase Systems	807
22.7	Database Backup and Recovery from Catastrophic Failures	808
22.8	Summary	809
	Review Questions	810
	Exercises	812
	Selected Bibliography	814

chapter **23** Emerging Database Technologies and Applications 817

23.1	Mobile Data Management	818
23.2	Multimedia Data Management	829

23.3 Geographic Information Systems (GIS)	840
23.4 Biological and Genomic Databases and Emerging Applications	854
Review Questions	867
Selected Bibliography	868
Credit	872

## chapter **24** Advanced Database Models and Applications    **873**

24.1 Active Database Concepts and Triggers	875
24.2 Temporal Database Concepts	885
24.3 Spatial Database Concepts	899
24.4 Multimedia Database Concepts	907
24.5 Introduction to Deductive Databases	912
24.6 Summary	925
Review Questions	926
Exercises	928
Selected Bibliography	931

## chapter **25** Introduction to Database Security    **935**

25.1 Introduction to Database Security Issues	936
25.2 Discretionary Access Control Based on Granting and Revoking Privileges	942
25.3 Mandatory Access Control and Role-Based Access Control for Multilevel Security	947
25.4 SQL Injection	955
25.5 Introduction to Statistical Database Security	959
25.6 Introduction to Flow Control	960
25.7 Encryption and Public Key Infrastructures	962
25.8 Privacy Issues and Preservation	966
25.9 Challenges of Database Security	967
25.10 Oracle Label-Based Security	968
25.11 Summary	970
Review Questions	972
Exercises	974
Selected Bibliography	974

## chapter **26** Introduction to Distributed Databases    **977**

26.1 Distributed Database Concepts	978
26.2 Types of Distributed Database Systems	983

26.3	Distributed Database Architectures	987
26.4	Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design	994
26.5	Query Processing and Optimization in Distributed Databases	1001
26.6	Overview of Transaction Management in Distributed Databases	1007
26.7	Overview of Concurrency Control and Recovery in Distributed Databases	1009
26.8	Distributed Catalog Management	1013
26.9	Current Trends in Distributed Databases	1014
26.10	Distributed Databases in Oracle	1015
26.11	Summary	1019
	Review Questions	1021
	Exercises	1023
	Selected Bibliography	1024

chapter **27** Introduction to Information Retrieval 1029

27.1	Information Retrieval (IR) Concepts	1030
27.2	Retrieval Models	1037
27.3	Types of Queries in IR Systems	1043
27.4	Text Preprocessing	1045
27.5	Inverted Indexing	1048
27.6	Evaluation Measures of Search Relevance	1050
27.7	Web Search and Analysis	1054
27.8	Trends in Information Retrieval	1064
27.9	Summary	1066
	Review Questions	1067
	Exercises	1069
	Selected Bibliography	1069

chapter **28** Data Mining Concepts 1071

28.1	Overview of Data Mining Technology	1072
28.2	Association Rules	1075
28.3	Classifications	1087
28.4	Clustering	1090
28.5	Approaches to Other Data Mining Problems	1093
28.6	Applications of Data Mining	1096
28.7	Commercial Data Mining Tools	1096
28.8	Summary	1099
	Review Questions	1099

Exercises	1100
Selected Bibliography	1101

## chapter **29** Overview of Data Warehousing and OLAP 1103

29.1 Introduction, Definition, and Terminology	1103
29.2 Characteristics of Data Warehouses	1105
29.3 Data Modeling for Data Warehouses	1106
29.4 Building a Data Warehouse	1111
29.5 Typical Functionality of a Data Warehouse	1114
29.6 Data Warehouse versus Views	1115
29.7 Difficulties of Implementing Data Warehouses	1116
29.8 Summary	1118
Review Questions	1118
Selected Bibliography	1119

## appendix **A** Alternative Diagrammatic Notations for ER Models 1121

## appendix **B** Parameters of Disks 1125

## appendix **C** Overview of the QBE Language 1129

C.1 Basic Retrievals in QBE	1129
C.2 Grouping, Aggregation, and Database Modification in QBE	1133

## appendix **D** Overview of the Hierarchical Data Model (located on the Companion Web site at <http://www.pearsoned.co.in/ramezelmasri>)

## appendix **E** Overview of the Network Data Model (located on the Companion Web site at <http://www.pearsoned.co.in/ramezelmasri>)

## Selected Bibliography 1137

## Index 1173