

Table of Contents

Preface xix

Acknowledgments xxiv

1. Introductory Database Concepts **1**

Chapter Objectives 1

- 1.1 Uses of Databases** 2
- 1.2 A Sample Database** 2
- 1.3 The Integrated Database Environment** 6
- 1.4 Roles in the Integrated Database Environment** 8
- 1.5 Advantages of the Integrated Database Approach** 11
- 1.6 Historical Developments in Information Systems** 13
 - 1.6.1 Big Data 18
- 1.7 Chapter Summary** 20

Exercises 21

2. Database Planning and Database Architecture **23**

Chapter Objectives 23

- 2.1 Data as a Resource** 24
- 2.2 Characteristics of Data** 24
 - 2.2.1 Data and Information 24
 - 2.2.2 Levels of Discussing Data 25
 - 2.2.3 Data Sublanguages 27

2.3	Stages in Database Design	27
2.4	Design Tools	31
2.4.1	Data Dictionary	31
2.4.2	Project Management Software	33
2.4.3	Diagramming Tools	33
2.4.4	CASE Packages	33
2.5	Database Administration	34
2.5.1	Planning and Design	34
2.5.2	Developing the Database	35
2.5.3	Database Management	36
2.6	The Three-Level Database Architecture	37
2.6.1	External Views	38
2.6.2	Logical and Conceptual Models	40
2.6.3	Internal Model	41
2.6.4	Data Independence	45
2.7	Overview of Data Models	46
2.7.1	The Entity-Relationship Model	46
2.7.2	Relational and Other Record-Based Models	48
2.7.3	Object-Oriented Model	49
2.7.4	Object-Relational Model	51
2.7.5	Large-Scale Data Models	51
2.8	Chapter Summary	53
	Exercises	54

3. The Entity-Relationship Model 57

Chapter Objectives 57

3.1	Purpose of the E-R Model	58
3.2	Entities	58
3.3	Attributes	60
3.3.1	Domains	60
3.3.2	Null Values	61
3.3.3	Multivalued Attributes	62
3.3.4	Composite Attributes	62
3.3.5	Derived Attributes	63

3.4	Keys	63
3.4.1	Superkeys	64
3.4.2	Candidate Keys	64
3.4.3	Primary Keys	65
3.5	Relationships	65
3.5.1	Degree of Relationships	66
3.5.2	Attributes of Relationship Sets	69
3.5.3	Cardinality of a Binary Relationship	69
3.5.4	Showing Cardinalities on an E-R Diagram	71
3.5.5	Participation Constraints	73
3.5.6	Using (min,max) Notation for Cardinality and Participation	74
3.6	Roles	75
3.7	Existence Dependency and Weak Entities	76
3.8	Sample E-R Diagrams	78
3.9	Rationale for Extending the E-R Model	85
3.10	Generalization and Specialization	85
3.10.1	Specialization	85
3.10.2	Generalization	88
3.10.3	Generalization Constraints—Disjointness, Completeness, Definition Method	88
3.10.4	Multiple Hierarchies and Inheritance	92
3.11	Union	93
3.12	Sample EE-R Diagrams	96
3.13	Chapter Summary	99
	Exercises	101

4. The Relational Model 107

Chapter Objectives 107

4.1	Advantages of the Relational Model	108
4.2	Relational Data Structures	108
4.2.1	Tables	109
4.2.2	Mathematical Relations	111
4.2.3	Database Relations and Tables	112

4.2.4	Properties of Relations	113
4.2.5	Degree and Cardinality	114
4.2.6	Relation Keys	114
4.3	Integrity Constraints: domain, key, foreign key, general constraints	116
4.4	Representing Relational Database Schemas	117
4.5	Relational Data Manipulation Languages	118
4.5.1	Categories of DMLs	118
4.5.2	Relational Algebra	119
4.6	Views	132
4.7	Mapping an E-R Model to a Relational Schema	133
4.8	Mapping an EE-R Model to a Relational Schema	140
4.8.1	Summary of E-R to Strictly Relational Mapping Concepts	140
4.8.2	Mapping EE-R Set Hierarchies to Relational Tables	142
4.8.3	Mapping Unions	143
4.8.4	EE-R to Relational Mapping Example	144
4.9	Chapter Summary	147
	Exercises	149

5. Relational Database Management Systems and SQL 153

Chapter Objectives 153

5.1	Brief History of SQL in Relational Database Systems	154
5.2	Architecture of a Relational Database Management System	155
5.3	Defining the Database: SQL DDL	156
5.3.1	CREATE DATABASE, CREATE SCHEMA	157
5.3.2	CREATE TABLE	158
5.3.3	CREATE INDEX	164
5.3.4	ALTER TABLE, RENAME TABLE	165
5.3.5	DROP Statements	166
5.3.6	Additional SQL DDL Example	167

5.4	Manipulating the Database: SQL DML	169
5.4.1	Introduction to the SELECT Statement	169
5.4.2	SELECT Using Multiple Tables	175
5.4.3	SELECT with Other Operators	186
5.4.4	Operators for Updating: UPDATE, INSERT, DELETE	196
5.4.5	Creating and Using Views	209
5.5	Active Databases	214
5.5.1	Enabling and Disabling Constraints	214
5.5.2	SQL Triggers	216
5.6	Using COMMIT and ROLLBACK Statements	222
5.7	Temporal Databases and SQL	222
5.7.1	User Valid Time	223
5.7.2	Transaction Time	228
5.8	SQL Programming	230
5.8.1	SQL PSMs	230
5.8.2	Embedded SQL	238
5.8.3	Dynamic SQL	243
5.8.4	Application Programming Interfaces (APIs)	243
5.9	JDBC	245
5.9.1	Developing a JDBC Application	245
5.9.2	The Statement Object	246
5.9.3	The PreparedStatement Object	248
5.9.4	The CallableStatement Object	250
5.10	The System Catalog	252
5.11	Chapter Summary	254
Exercises		255

6. Normalization and Denormalization 259

Chapter Objectives	259
6.1 Objectives of Normalization	260
6.2 Insertion, Update, and Deletion Anomalies	260
6.3 Functional Dependency	262
6.4 Superkeys, Candidate Keys, and Primary Keys	265

6.5	Normalization Using Primary Keys	266
6.5.1	First Normal Form	266
6.5.2	Full Functional Dependency and Second Normal Form	269
6.5.3	Transitive Dependency and Third Normal Form	272
6.5.4	Boyce-Codd Normal Form	274
6.5.5	Comprehensive Example of Functional Dependencies	277
6.6	Properties of Relational Decompositions	281
6.6.1	Attribute Preservation	281
6.6.2	Dependency Preservation	281
6.6.3	Lossless Decomposition	282
6.6.4	Decomposition Algorithm for Boyce-Codd Normal Form with Lossless Join	284
6.7	The Normalization Process	285
6.7.1	Analysis	285
6.7.2	Synthesis	285
6.7.3	Normalization from an Entity-Relationship Diagram	286
6.8	When to Stop Normalizing	286
6.9	Non-normalized Databases	287
6.10	Chapter Summary	288
	Exercises	290

7. Object-Based Models **295**

Chapter Objectives	295	
7.1	Rationale for the Object-Oriented Data Model	296
7.2	Object-Oriented Data Concepts	297
7.2.1	Objects and Literals	297
7.2.2	Classes	298
7.2.3	Class Hierarchies and Inheritance	300
7.2.4	Object Identity	304
7.3	Object-Oriented Data Modeling Using UML	305
7.4	The ODMG Model and ODL	308
7.4.1	Class Declarations	308
7.4.2	Extent	308

7.4.3	Attributes	308
7.4.4	Relationships	313
7.4.5	Methods	314
7.4.6	Classes and Inheritance	314
7.4.7	N-ary Relationships and M:N Relationships with Attributes	315
7.4.8	Keys	316
7.5	Object Query Language	316
7.6	Developing an OO Database	321
7.7	Developing an OO Database in Caché	322
7.8	Extending the Relational Model	328
7.8.1	Additional Fundamental Data Types	328
7.8.2	Collection Types	329
7.8.3	User-Defined Data Types (UDTs)	332
7.8.4	Reference Types	337
7.8.5	Type Hierarchies in Standard SQL	341
7.8.6	Type Hierarchies in Oracle	342
7.8.7	Nested Tables in Oracle	344
7.8.8	Oracle Object Views	347
7.9	Converting a UML Diagram to an Object-Relational Database Model	348
7.10	Converting an EE-R Diagram to an Object-Relational Database Model	355
7.11	Chapter Summary	357
	Exercises	359

8. Introduction to Database Security 361

Chapter Objectives 361

8.1	Issues in Database Security	362
8.1.1	Accidental Security Threats	363
8.1.2	Deliberate Security Threats	364
8.2	Fundamentals of Access Control	365
8.2.1	Physical Security	365
8.2.2	Information System Access Control	365

8.3	Database Access Control	367
8.4	Using Views for Access Control	368
8.5	Security Logs and Audit Trails	369
8.6	Encryption	370
8.6.1	Symmetric Key Encryption	371
8.6.2	Public-Key Encryption	372
8.7	SQL Data Control Language	373
8.8	Security in Oracle	377
8.8.1	Security Features	377
8.8.2	Administrative Accounts	378
8.8.3	Security Tools	379
8.9	Statistical Database Security	381
8.10	SQL Injection	382
8.10.1	Examples of SQL Injection	383
8.10.2	Mitigation of SQL Injection	385
8.11	Database Security and the Internet	385
8.11.1	Proxy Servers	386
8.11.2	Firewalls	386
8.11.3	Digital Signatures	386
8.11.4	Certification Authorities	387
8.12	Chapter Summary	388
	Exercises	389

9. Transaction Management **393**

Chapter Objectives 393

9.1	ACID Properties of Transactions	394
9.2	Need for Concurrency Control	397
9.2.1	The Lost Update Problem	398
9.2.2	The Uncommitted Update Problem	399
9.2.3	The Problem of Inconsistent Analysis	401
9.3	Serializability	402

9.4	Locking	406
9.4.1	Deadlock	407
9.4.2	Two-Phase Locking Protocol	411
9.4.3	Levels of Locking	412
9.5	Timestamping	415
9.5.1	Basic Timestamping Protocol	416
9.5.2	Thomas's Write Rule	418
9.5.3	Multiversion Timestamping	418
9.6	Optimistic Techniques	420
9.7	Need for Recovery	422
9.8	Recovery Techniques	423
9.8.1	Deferred Update Protocol	424
9.8.2	Checkpoints	426
9.8.3	Immediate Update Protocol	427
9.8.4	Shadow Paging	427
9.8.5	Overview of the ARIES Recovery Algorithm	428
9.9	Transaction Management in Oracle	429
9.10	Chapter Summary	433
Exercises		435

10. Distributed Databases 439

Chapter Objectives 439

10.1	Rationale for Distribution	440
10.2	Architectures for a Distributed System	440
10.2.1	Distributed Processing Using a Centralized Database	441
10.2.2	Client-Server Systems	441
10.2.3	Parallel Databases	442
10.2.4	Distributed Databases	446
10.3	Components of a Distributed Database System	447
10.4	Data Placement	449
10.5	Transparency	454

10.6 Transaction Control for Distributed Databases 456

10.6.1 Concurrency Control 457

10.6.2 Recovery 462

10.7 Distributed Query Processing 466

10.7.1 Steps in Distributed Query Processing 466

10.7.2 The Semijoin Operation 469

10.8 Chapter Summary 470**Exercises 472****11. Data Warehouses and Data Mining 477****Chapter Objectives 477****11.1 Origins of Data Warehousing and Data Mining 478****11.2 Operational Databases and Data Warehouses 480****11.3 Architecture of a Data Warehouse 481****11.4 Developing a Data Warehouse 482**

11.4.1 Top-Down Method 482

11.4.2 Bottom-Up Method 483

11.4.3 Comparison of Methods 485

11.5 Data Models for Data Warehouses 485

11.5.1 ROLAP Schemas 486

11.5.2 Optimization and Index Techniques for ROLAP 488

11.5.3 MOLAP Data Cubes and Hypercubes 490

11.6 Data Warehouse Queries and SQL 493

11.6.1 Aggregate Functions 494

11.6.2 RANK Functions 494

11.6.3 CUBE and ROLLUP Functions 495

11.6.4 Analytic Functions 496

11.7 Views and View Materialization 496**11.8 Data Warehouse Vendors 503****11.9 Data Mining 504****11.10 Purpose of Data Mining 505**

11.11	Types of Knowledge Discovered	506
11.12	Models and Methods Used	508
11.12.1	Data Mining Process Model	508
11.12.2	Regression	510
11.12.3	Decision Trees	511
11.12.4	Neural Networks	512
11.12.5	Clustering	514
11.12.6	Genetic Algorithms	515
11.13	Applications of Data Mining	515
11.13.1	Retailing	515
11.13.2	Banking and Finance	516
11.13.3	Manufacturing	516
11.13.4	Science and Medicine	516
11.13.5	Homeland Security	517
11.13.6	Search Engines	518
11.14	Data Mining Software Providers	518
11.15	Chapter Summary	519
Exercises		521

12. Big Data and NoSQL 525

Chapter Objectives	525	
12.1	Introduction	526
12.2	Defining Big Data	529
12.3	Hadoop	531
12.3.1	Hadoop Distributed File System	531
12.3.2	MapReduce	533
12.3.3	Querying Data in Hadoop	537
12.3.4	Hive and HiveQL	538
12.3.5	Hadoop Software Products	550
12.4	NoSQL	550
12.4.1	Defining NoSQL Technology	551
12.4.2	The HBase Column-Oriented Datastore	554
12.4.3	NoSQL Software Products	563

12.5	NewSQL	563
12.6	Chapter Summary	564
	Exercises	566

13. Relational Query Optimization 569

Chapter Objectives		569
13.1	Query Processing and Query Optimization	570
13.2	Algebraic Techniques for Query Transformation	572
13.2.1	The Query Tree	572
13.2.2	An SQL Query and Its Relational Algebra Translation	573
13.2.3	Performing SELECT Operations Early	574
13.2.4	Evaluating Conjunctive Conditions	575
13.2.5	Performing PROJECT Early	578
13.2.6	Equivalence of Algebraic Operations	578
13.2.7	Heuristics for Query Optimization	582
13.3	Processing Techniques and Cost Estimation	583
13.3.1	Cost Factors	583
13.3.2	Cost of Processing Selects	585
13.3.3	Processing Joins	588
13.3.4	Processing Other Operations	593
13.4	Pipelining	596
13.5	Query Optimization in Oracle	597
13.6	Chapter Summary	599
Exercises		600

14. Databases and the Internet 605

Chapter Objectives	605	
14.1	Introduction	606

14.2 Fundamental Concepts of the Internet and the World Wide Web 606

- 14.2.1 Browsers, Links, URIs, and URLs 607
- 14.2.2 HTTP 607
- 14.2.3 HTML 609

14.3 Extensible Markup Language (XML) 611

- 14.3.1 Standalone XML Documents 612
- 14.3.2 DTDs 614
- 14.3.3 XML Schema 617

14.4 Tiered Architectures 619

- 14.4.1 Single-Tier Architecture 619
- 14.4.2 Two-Tier Architecture 619
- 14.4.3 Three-Tier Architecture 623

14.5 Web Programming with Oracle 628

- 14.5.1 PL/SQL Web Toolkit 628
- 14.5.2 PL/SQL Server Pages (PSP) 632

14.6 The Semi-Structured Data Model 634

- 14.6.1 Graphical Representation 634
- 14.6.2 XML Parsers 637
- 14.6.3 XML Data Manipulation 638

14.7 XML and Relational Databases 644**14.8 XML in Oracle 645**

- 14.8.1 XML DB 645
- 14.8.2 Oracle XML Developer Kits 652

14.9 Chapter Summary 652**Exercises 655****15. Social and Ethical Issues 657****Chapter Objectives 657****15.1 Computerization and Society 658**

- 15.1.1 Ethical Issues in Computing 658
- 15.1.2 Ethical Standards for Computing Professionals 659

15.2 Intellectual Property 660

- 15.2.1 Definition of Intellectual Property 662
- 15.2.2 Legal Protections for Intellectual Property 663
- 15.2.3 Intellectual Property Protection for Software 675

15.3 Databases and Privacy Issues 677

- 15.3.1 Privacy and Security 678
- 15.3.2 Privacy as a Human Right 679

15.4 Human Factors 687

- 15.4.1 Human Factors in Software Development 687
- 15.4.2 The Human-Database Interface 689
- 15.4.3 Usability Testing for Database Applications 690

15.5 Chapter Summary 691**Exercises 693****Index 695**