

INTERNATIONAL EDITION

INTRODUCTORY VERSION

Database Systems

An Application-Oriented Approach

SECOND EDITION



Michael KIFER Arthur BERNSTEIN Philip M. LEWIS

Contents

Preface	xvii
PART ONE Introduction	1
1 Overview of Databases and Transactions	3
1.1 What Are Databases and Transactions?	3
1.2 Features of Modern Database and Transaction Processing Systems	6
1.3 Major Players in the Implementation and Support of Database and Transaction Processing Systems	7
1.4 Decision Support Systems—OLAP and OLTP	9
2 The Big Picture	13
⑧ 2.1 Case Study: A Student Registration System	13
2.2 Introduction to Relational Databases	14
2.3 What Makes a Program a Transaction—The ACID Properties	20
Bibliographic Notes	25
Exercises	25
PART TWO Database Management	29
3 The Relational Data Model	31
3.1 What Is a Data Model?	31
3.2 The Relational Model	35
3.2.1 Basic Concepts	35
3.2.2 Integrity Constraints	38
3.3 SQL—Data Definition Sublanguage	46
3.3.1 Specifying the Relation Type	46
3.3.2 The System Catalog	46
3.3.3 Key Constraints	47
3.3.4 Dealing with Missing Information	48

3.3.5	Semantic Constraints	49
3.3.6	User-Defined Domains	53
3.3.7	Foreign-Key Constraints	53
3.3.8	Reactive Constraints	56
3.3.9	Database Views	59
3.3.10	Modifying Existing Definitions	60
3.3.11	SQL-Schemas	62
3.3.12	Access Control	63
	Bibliographic Notes	65
	Exercises	66

4	Conceptual Modeling of Databases with Entity-Relationship Diagrams and the Unified Modeling Language	69
4.1	Conceptual Modeling with the E-R Approach	70
4.2	Entities and Entity Types	70
4.3	Relationships and Relationship Types	73
4.4	Advanced Features in Conceptual Data Modeling	78
4.4.1	Entity Type Hierarchies	78
4.4.2	Participation Constraints	81
4.4.3	The Part-of Relationship	83
4.5	From E-R Diagrams to Relational Database Schemas	86
4.5.1	Representation of Entities	86
4.5.2	Representation of Relationships	88
4.5.3	Representing IsA Hierarchies in the Relational Model	90
4.5.4	Representation of Participation Constraints	92
4.5.5	Representation of the Part-of Relationship	94
4.6	UML: A New Kid on the Block*	95
4.6.1	Representing Entities in UML	96
4.6.2	Representing Relationships in UML	97
4.6.3	Advanced Modeling Concepts in UML	101
4.6.4	Translation to SQL	105
4.7	A Brokerage Firm Example	106
4.7.1	An Entity-Relationship Design	106
4.7.2	A UML Design*	110
④ 4.8	Case Study: A Database Design for the Student Registration System	111
4.8.1	The Database Part of the Requirements Document	112
4.8.2	The Database Design	113
4.9	Limitations of Data Modeling Methodologies	119
	Bibliographic Notes	123
	Exercises	123

5	Relational Algebra and SQL	127
5.1	Relational Algebra: Under the Hood of SQL	128
5.1.1	Basic Operators	128
5.1.2	Derived Operators	137
5.2	The Query Sublanguage of SQL	147
5.2.1	Simple SQL Queries	148
5.2.2	Set Operations	154
5.2.3	Nested Queries	157
5.2.4	Quantified Predicates	163
5.2.5	Aggregation over Data	164
5.2.6	Join Expressions in the FROM Clause	170
5.2.7	A Simple Query Evaluation Algorithm	171
5.2.8	More on Views in SQL	174
5.2.9	Materialized Views	177
5.2.10	The Null Value Quandary	181
5.3	Modifying Relation Instances in SQL	182
5.3.1	Inserting Data	182
5.3.2	Deleting Data	184
5.3.3	Updating Existing Data	185
5.3.4	Updates on Views	185
	Bibliographic Notes	187
	Exercises	188
6	Database Design with the Relational Normalization Theory	193
6.1	The Problem of Redundancy	193
6.2	Decompositions	195
6.3	Functional Dependencies	198
6.4	Properties of Functional Dependencies	200
6.5	Normal Forms	207
6.5.1	The Boyce-Codd Normal Form	208
6.5.2	The Third Normal Form	210
6.6	Properties of Decompositions	211
6.6.1	Lossless and Lossy Decompositions	212
6.6.2	Dependency-Preserving Decompositions	215
6.7	An Algorithm for BCNF Decomposition	219
6.8	Synthesis of 3NF Schemas	221
6.8.1	Minimal Cover	222
6.8.2	3NF Decomposition through Schema Synthesis	224
6.8.3	BCNF Decomposition through 3NF Synthesis	226
6.9	The Fourth Normal Form	228

6.10	Advanced 4NF Design *	233
6.10.1	MVDs and Their Properties	234
6.10.2	The Difficulty of Designing for 4NF	235
6.10.3	A 4NF Decomposition How-To	238
6.11	Summary of Normal Form Decomposition	240
⑧ 6.12	Case Study: Schema Refinement for the Student Registration System	241
6.13	Tuning Issues: To Decompose or Not to Decompose?	244
	Bibliographic Notes	245
	Exercises	246
7	Triggers and Active Databases	251
7.1	What Is a Trigger?	251
7.2	Semantic Issues in Trigger Handling	252
7.3	Triggers in SQL:1999	256
7.4	Avoiding a Chain Reaction	264
	Bibliographic Notes	265
	Exercises	265
8	Using SQL in an Application	267
8.1	What Are the Issues Involved?	267
8.2	Embedded SQL	268
8.2.1	Status Processing	271
8.2.2	Sessions, Connections, and Transactions	273
8.2.3	Executing Transactions	274
8.2.4	Cursors	276
8.2.5	Stored Procedures on the Server	282
8.3	More on Integrity Constraints	285
8.4	Dynamic SQL	286
8.4.1	Statement Preparation in Dynamic SQL	287
8.4.2	Prepared Statements and the Descriptor Area *	290
8.4.3	Cursors	293
8.4.4	Stored Procedures on the Server	293
8.5	JDBC and SQLJ	294
8.5.1	JDBC Basics	294
8.5.2	Prepared Statements	297
8.5.3	Result Sets and Cursors	297
8.5.4	Obtaining Information about a Result Set	300
8.5.5	Status Processing	300

8.5.6	Executing Transactions	301
8.5.7	Stored Procedures on the Server	302
8.5.8	An Example	303
8.5.9	SQLJ: Statement-Level Interface to Java	303
8.6	ODBC*	307
8.6.1	Prepared Statements	309
8.6.2	Cursors	309
8.6.3	Status Processing	312
8.6.4	Executing Transactions	312
8.6.5	Stored Procedures on the Server	313
8.6.6	An Example	313
8.7	Comparison	315
	Bibliographic Notes	316
	Exercises	316

PART THREE Optimizing DBMS Performance and Transaction Processing **319**

9 Physical Data Organization and Indexing **321**

9.1	Disk Organization	322
9.1.1	RAID Systems	326
9.2	Heap Files	329
9.3	Sorted Files	333
9.4	Indices	337
9.4.1	Clustered versus Unclustered Indices	340
9.4.2	Sparse versus Dense Indices	342
9.4.3	Search Keys Containing Multiple Attributes	344
9.5	Multilevel Indexing	347
9.5.1	Index-Sequential Access	350
9.5.2	B ⁺ Trees	353
9.6	Hash Indexing	360
9.6.1	Static Hashing	360
9.6.2	Dynamic Hashing Algorithms	363
9.7	Special-Purpose Indices	371
9.7.1	Bitmap Indices	371
9.7.2	Join Indices	372
9.8	Tuning Issues: Choosing Indices for an Application	373
	Bibliographic Notes	374
	Exercises	375

10 The Basics of Query Processing	379
10.1 Overview of Query Processing	379
10.2 External Sorting	380
10.3 Computing Projection, Union, and Set Difference	384
10.4 Computing Selection	386
10.4.1 Selections with Simple Conditions	387
10.4.2 Access Paths	389
10.4.3 Selections with Complex Conditions	391
10.5 Computing Joins	392
10.5.1 Computing Joins Using Simple Nested Loops	393
10.5.2 Sort-Merge Join	396
10.5.3 Hash Join	398
10.6 Multirelational Joins*	399
10.7 Computing Aggregate Functions	401
Bibliographic Notes	401
Exercises	401
11 An Overview of Query Optimization	405
11.1 Query Processing Architecture	405
11.2 Heuristic Optimization Based on Algebraic Equivalences	407
11.3 Estimating the Cost of a Query Execution Plan	410
11.4 Estimating the Size of the Output	418
11.5 Choosing a Plan	420
Bibliographic Notes	425
Exercises	425
12 Database Tuning	429
12.1 Disk Caches	430
12.1.1 Tuning the Cache	431
12.2 Tuning the Schema	433
12.2.1 Indices	433
12.2.2 Denormalization	440
12.2.3 Repeating Groups	441
12.2.4 Partitioning	442
12.3 Tuning the Data Manipulation Language	443
12.4 Tools	446
12.5 Managing Physical Resources	447
12.6 Influencing the Optimizer	448
Bibliographic Notes	451

13 An Overview of Transaction Processing **455**

- 13.1 Isolation 455**
 - 13.1.1 Serializability 456
 - 13.1.2 Two-Phase Locking 458
 - 13.1.3 Deadlock 462
 - 13.1.4 Locking in Relational Databases 463
 - 13.1.5 Isolation Levels 465
 - 13.1.6 Lock Granularity and Intention Locks 468
 - 13.1.7 Summary 471
- 13.2 Atomicity and Durability 472**
 - 13.2.1 The Write-Ahead Log 472
 - 13.2.2 Recovery from Mass Storage Failure 476
- 13.3 Implementing Distributed Transactions 477**
 - 13.3.1 Atomicity and Durability—The Two-Phase Commit Protocol 478
 - 13.3.2 Global Serializability and Deadlock 480
 - 13.3.3 Replication 482
 - 13.3.4 Summary 484
 - Bibliographic Notes 484
 - Exercises 485

PART FOUR Software Engineering Issues and Documentation **487**

14 Requirements and Specifications **489**

- 14.1 Software Engineering Methodology 489**
 - 14.1.1 UML Use Cases 490
- ④ **14.2 The Requirements Document for the Student Registration System 493**
- ④ **14.3 Requirements Analysis—New Issues 500**
- ④ **14.4 Specifying the Student Registration System 502**
 - 14.4.1 UML Sequence Diagrams 503
- ④ **14.5 The Specification Document for the Student Registration System: Section III 504**
- 14.6 The Next Step in the Software Engineering Process 506**
 - Bibliographic Notes 506
 - Exercises 507

15	Design, Coding, and Testing	509
15.1	The Design Process	509
15.1.1	Database Design	510
15.1.2	Describing the Behavior of Objects with UML State Diagrams	510
15.1.3	Structure of the Design Document	512
15.1.4	Design Review	514
15.2	Test Plan	515
15.3	Project Planning	518
15.4	Coding	521
15.5	Incremental Development	523
15.6	The Project Management Plan	524
⑤ 15.7	Design and Code for the Student Registration System	525
15.7.1	Completing the Database Design: Integrity Constraints	526
15.7.2	Design of the Registration Transaction	528
15.7.3	Partial Code for the Registration Transaction	530
	Bibliographic Notes	533
	Exercises	533
PART FIVE	Advanced Topics in Databases	535
16	Introduction to Object Databases	537
16.1	Shortcomings of the Relational Data Model	537
16.2	The Conceptual Object Data Model	543
16.2.1	Objects and Values	544
16.2.2	Classes	545
16.2.3	Types	546
16.2.4	Object-Relational Databases	549
16.3	Objects in SQL:1999 and SQL:2003	550
16.3.1	Row Types	551
16.3.2	User-Defined Types	552
16.3.3	Objects	553
16.3.4	Querying User-Defined Types	554
16.3.5	Updating User-Defined Types	555
16.3.6	Reference Types	558
16.3.7	Inheritance	560
16.3.8	Collection Types	561
	Bibliographic Notes	563
	Exercises	564

17	Introduction to XML and Web Data	567
17.1	Semistructured Data	567
17.2	Overview of XML	570
17.2.1	XML Elements and Database Objects	573
17.2.2	XML Attributes	575
17.2.3	Namespaces	577
17.2.4	Document Type Definitions	582
17.2.5	Inadequacy of DTDs as a Data Definition Language	585
17.3	XML Schema	586
17.3.1	XML Schema and Namespaces	587
17.3.2	Simple Types	590
17.3.3	Complex Types	595
17.3.4	Putting It Together	603
17.3.5	Shortcuts: Anonymous Types and Element References	606
17.3.6	Integrity Constraints	608
17.4	XML Query Languages	615
17.4.1	XPath: A Lightweight XML Query Language	616
17.4.2	SQL/XML	623
	Bibliographic Notes	633
	Exercises	634
	Bibliography	639
	Index	649