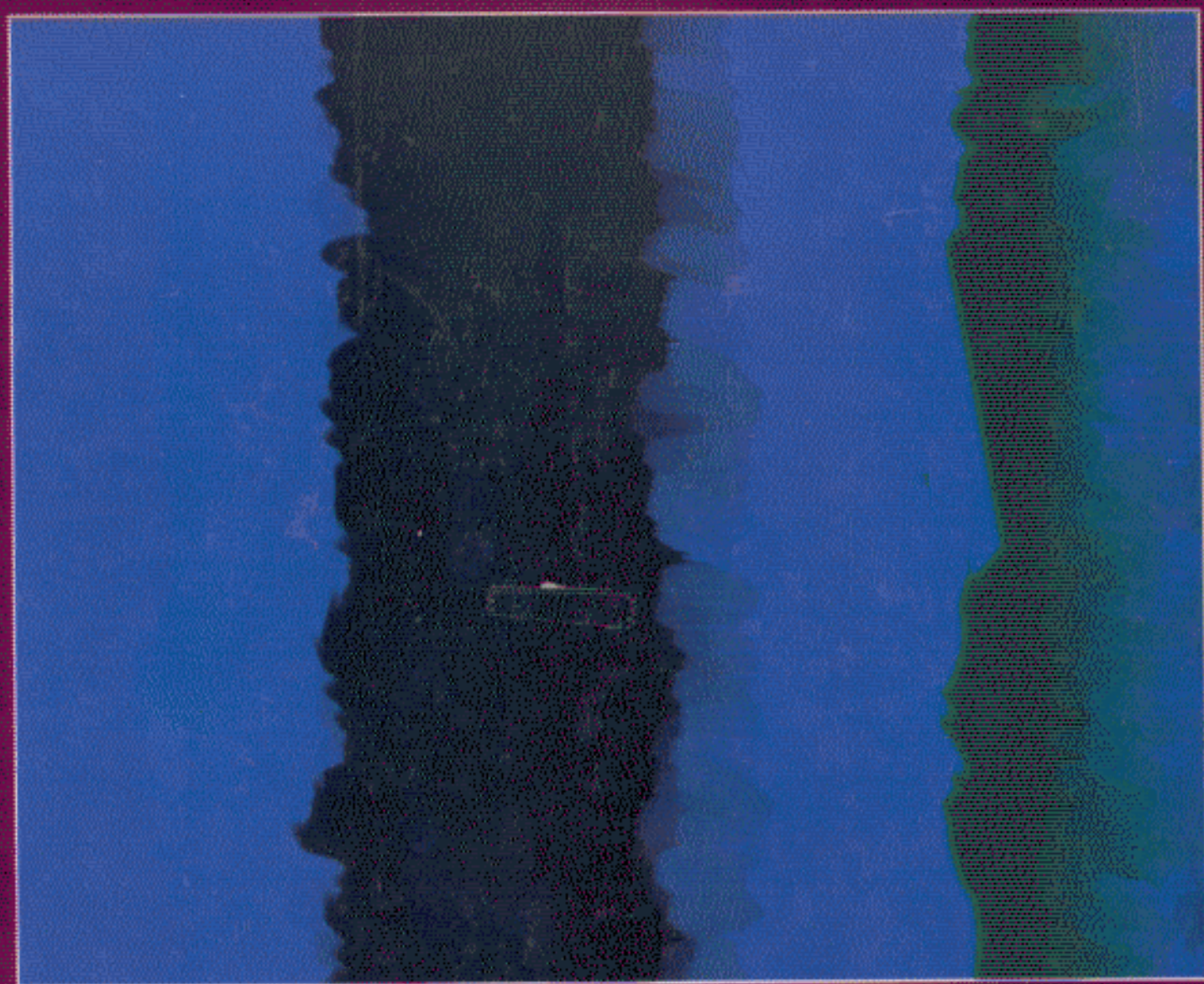


MODERN DATABASE MANAGEMENT

SIXTH EDITION



JEFFREY A. HOFFER • MARY B. PRESCOTT
FRED R. McFADDEN

Preface

XXV

Part

The Context of Database Management

PART I OVERVIEW 2

1 The Database Environment 3

Learning Objectives 3

Introduction 3

Basic Concepts and Definitions 4

Data 4

Data Versus Information 5

Metadata 6

Traditional File Processing Systems 7

FILE PROCESSING SYSTEMS AT PINE VALLEY FURNITURE COMPANY 7

Disadvantages of File Processing Systems 8

Program-Data Dependence 8

Duplication of Data 8

Limited Data Sharing 9

Lengthy Development Times 9

Excessive Program Maintenance 9

The Database Approach 9

THE DATABASE APPROACH AT PINE VALLEY FURNITURE COMPANY 9

Enterprise Data Model 10

Relational Databases 11

Implementing Relational Databases 11

A Database Application 13

The Range of Database Applications 14

Personal Databases 14

Workgroup Databases 16

Department Databases 17

Enterprise Databases 18

Internet, Intranet, and Extranet Databases 19

Summary of Database Applications 20

Advantages of the Database Approach 21

Program-Data Independence 21

Minimal Data Redundancy 21

Improved Data Consistency	21
Improved Data Sharing	21
Increased Productivity of Application Development	21
Enforcement of Standards	22
Improved Data Quality	22
Improved Data Accessibility and Responsiveness	22
Reduced Program Maintenance	22
Cautions About Database Benefits	23
Costs and Risks of the Database Approach	23
New, Specialized Personnel	23
Installation and Management Cost and Complexity	23
Conversion Costs	24
Need for Explicit Backup and Recovery	24
Organizational Conflict	24
Components of the Database Environment	24
Evolution of Database Systems	26
1960s	26
1970s	27
1980s	27
1990s	27
2000 and Beyond	27
Summary	28
Chapter Review	29
Key Terms	29
Review Questions	29
Problems and Exercises	30
Field Exercises	30
References	31
Further Reading	31
Web Resources	31
Project Case: Mountain View Community Hospital	32

2 Database Development Process

34

Learning Objectives	34
Introduction	34
Database Development Within Information Systems Development	35
Information Systems Architecture	36
Information Engineering	36
Information Systems Planning	37
Identifying Strategic Planning Factors	37
Identifying Corporate Planning Objects	37
Developing an Enterprise Model	38
Database Development Process	40
Systems Development Life Cycle	41
Enterprise Modeling	42
Conceptual Data Modeling	42
Logical Database Design	43
Physical Database Design and Definition	43
Database Implementation	43
Database Maintenance	44
Alternative IS Development Approaches	44
The Role of CASE and a Repository	45

Managing the People Involved in Database Development	46
Three-Schema Architecture for Database Development	49
Three-Tiered Database Location Architecture	51
DEVELOPING A DATABASE APPLICATION FOR PINE VALLEY FURNITURE	53
Matching User Needs to the Information Systems Architecture	54
Analyzing Database Requirements	56
Designing the Database	57
Using the Database	60
Administering the Database	62
Summary	62
Chapter Review	63
Key Terms	63
Review Questions	63
Problems and Exercises	64
Field Exercises	66
References	66
Further Reading	66
Web Resources	66
Project Case: Mountain View Community Hospital	67

Part

Database Analysis

PART II OVERVIEW 74

3 Modeling Data in the Organization

75

Learning Objectives 75

Introduction 75

Modeling the Rules of the Organization 77

 Overview of Business Rules 77

 The Business Rules Paradigm 78

 Scope of Business Rules 78

 Good Business Rules 78

 Gathering Business Rules 79

 Data Names and Definitions 79

 Data Names 79

 Data Definitions 80

 Good Data Definitions 81

The E-R Model 82

 Sample E-R Diagram 82

 E-R Model Notation 84

Entity-Relationship Model Constructs 84

 Entities 85

 Entity Type Versus Entity Instance 86

 Entity Type Versus System Input, Output, or User 86

 Strong Versus Weak Entity Types 87

 Naming and Defining Entity Types 88

 Attributes 90

 Simple Versus Composite Attributes 91

Single-Valued Versus Multivalued Attributes	91
Stored Versus Derived Attributes	91
Identifier Attribute	92
Naming and Defining Attributes	93
Relationships	95
Basic Concepts and Definitions in Relationships	97
Attributes on Relationships	97
Associative Entities	97
Degree of a Relationship	99
Unary Relationship	100
Binary Relationship	100
Ternary Relationship	100
Attributes or Relationship?	103
Cardinality Constraints	103
Minimum Cardinality	103
Maximum Cardinality	103
Some Examples	105
A Ternary Relationship	106
Modeling Time-Dependent Data	106
Multiple Relationships	109
Naming and Defining Relationships	110
E-R MODELING EXAMPLE: PINE VALLEY FURNITURE COMPANY	112
DATABASE PROCESSING AT PINE VALLEY FURNITURE	114
Showing Product Information	114
Showing Customer Information	115
Showing Customer Order Status	115
Showing Product Sales	117
Summary	117
Chapter Review	118
Key Terms	118
Review Questions	119
Problems and Exercises	120
Field Exercises	122
References	123
Further Reading	123
Web Resources	124
Project Case: Mountain View Community Hospital	125

4 The Enhanced E-R Model and Business Rules

127

Learning Objectives	127
Introduction	127
Representing Supertypes and Subtypes	128
Basic Concepts and Notation	129
An Example	129
Attribute Inheritance	131
When to Use Supertype/Subtype Relationships	131
Representing Specialization and Generalization	132
Generalization	132
Specialization	132
Combining Specialization and Generalization	135
Specifying Constraints in Supertype/Subtype Relationships	135
Specifying Completeness Constraints	135

Total Specialization Rule	135
Partial Specialization Rule	*135
Specifying Disjointness Constraints	137
Disjoint Rule	137
Overlap Rule	137
Defining Subtype Discriminators	137
Disjoint Subtypes	137
Overlapping Subtypes	139
Defining Supertype/Subtype Hierarchies	140
An Example	140
Summary of Supertype/Subtype Hierarchies	142
EER MODELING EXAMPLE: PINE VALLEY FURNITURE	142
Entity Clustering	145
Business Rules Revisited	148
Classification of Business Rules	149
Stating a Structural Assertion	151
Derived Facts	151
Stating an Action Assertion	151
Types of Action Assertions	152
Representing and Enforcing Business Rules	153
Sample Business Rules	153
Identifying and Testing Business Rules	155
Summary	156
Chapter Review	157
Key Terms	157
Review Questions	157
Problems and Exercises	157
Field Exercises	159
References	159
Further Reading	159
Web Resources	159
Project Case: Mountain View Community Hospital	160

Part

Database Design

PART III OVERVIEW 164

5 Logical Database Design and the Relational Model

165

Learning Objectives	165
Introduction	165
The Relational Data Model	166
Basic Definitions	166
Relational Data Structure	166
Relational Keys	167
Properties of Relations	168
Removing Multivalued Attributes from Tables	168
Example Database	169
Integrity Constraints	170
Domain Constraints	171
Entity Integrity	171
Referential Integrity	171

Action Assertions	173
Creating Relational Tables	173
Well-Structured Relations	174
Transforming EER Diagrams into Relations	175
Step 1: Map Regular Entities	175
Composite Attributes	176
Multivalued Attributes	176
Step 2: Map Weak Entities	178
Step 3: Map Binary Relationships	179
Map Binary One-to-Many Relationships	179
Map Binary Many-to-Many Relationships	180
Map Binary One-to-One Relationships	180
Step 4: Map Associative Entities	182
Identifier Not Assigned	182
Identifier Assigned	183
Step 5: Map Unary Relationships	184
Unary One-to-Many Relationships	184
Unary Many-to-Many Relationships	185
Step 6: Map Ternary (and n -ary) Relationships	186
Step 7: Map Supertype/Subtype Relationships	186
Introduction to Normalization	189
Steps in Normalization	189
Functional Dependencies and Keys	190
Determinants	191
Candidate Keys	191
The Basic Normal Forms	192
First Normal Form	192
Second Normal Form	193
Third Normal Form	193
Normalizing Summary Data	195
Merging Relations	196
An Example	197
View Integration Problems	197
Synonyms	197
Homonyms	197
Transitive Dependencies	198
Supertype/Subtype Relationships	198
A Final Step for Defining Relational Keys	199
Summary	201
Chapter Review	202
Key Terms	202
Review Questions	202
Problems and Exercises	203
Field Exercises	205
References	206
Further Reading	206
Web Resources	206
Project Case: Mountain View Community Hospital	207

Introduction	208
Physical Database Design Process	209
Data Volume and Usage Analysis	210
Designing Fields	211
Choosing Data Types	212
Coding and Compression Techniques	213
Controlling Data Integrity	213
Handling Missing Data	214
Designing Physical Records and Denormalization	215
Denormalization	215
Designing Physical Files	221
Pointer	222
File Organizations	222
Sequential File Organizations	223
Indexed File Organizations	223
Hashed File Organizations	229
Summary of File Organizations	230
Clustering Files	230
Designing Controls for Files	231
Using and Selecting Indexes	232
Creating a Primary Key Index	232
Creating a Secondary (Nonunique) Key Index	232
When to Use Indexes	233
RAID: Improving File Access Performance by Parallel Processing	234
Choosing Among RAID Levels	235
RAID-0	235
RAID-1	235
RAID 0+1	238
RAID-2	238
RAID-3	238
RAID-4	238
RAID-5	239
RAID Performance	239
Designing Databases	239
Choosing Database Architectures	239
Optimizing for Query Performance	242
Parallel Query Processing	243
Overriding Automatic Query Optimization	243
Picking Data Block Size	244
Balancing I/O Across Disk Controllers	245
Guidelines for Better Query Design	245
Summary	247
Chapter Review	248
Key Terms	248
Review Questions	248
Problems and Exercises	249
Field Exercises	250
References	251
Further Reading	251
Web Resources	251
Project Case: Mountain View Community Hospital	252

7 SQL**257**

- Learning Objectives 257
- Introduction 257
- History of the SQL Standard 258
- The Role of SQL in a Database Architecture 259
- The SQL Environment 261
- Defining a Database in SQL 265
 - Generating SQL Database Definitions 265
 - Creating Tables 266
 - Using and Defining Views 268
 - Materialized Views 271
 - Creating Data Integrity Controls 271
 - Changing Table Definitions 273
 - Removing Tables 273
- Inserting, Updating, and Deleting Data 273
 - Batch Input 274
 - Deleting Database Contents 274
 - Changing Database Contents 275
- Internal Schema Definitions in RDBMSs 275
 - Creating Indexes 276
- Processing Single Tables 276
 - Clauses of the SELECT Statement 277
 - Using Expressions 278
 - Using Functions 279
 - Using Wildcards 280
 - Comparison Operators 280
 - Using Boolean Operators 281
 - Ranges 282
 - Distinct 283
 - IN and NOT IN Lists 284
 - Sorting Results: The ORDER BY Clause 285
 - Categorizing Results: The GROUP BY Clause 286
 - Qualifying Results by Categories: The HAVING Clause 287
- Summary 288
- Chapter Review 289
 - Key Terms 289
 - Review Questions 289
 - Problems and Exercises 290
 - Field Exercises 292
 - References 292
 - Further Reading 292
 - Web Resources 293
- Project Case: Mountain View Community Hospital 294**

8 Advanced SQL**295**

- Learning Objectives 295

Introduction	295
Processing Multiple Tables	296
Equi-Join	296
Natural Join	297
Outer Join	298
Union Join	299
Sample Multiple Join Involving Four Tables	300
Subqueries	300
Correlated Subqueries	303
Using Derived Tables	305
Combining Queries	306
Conditional Expressions	306
Ensuring Transaction Integrity	307
Data Dictionary Facilities	308
SQL-99 Enhancements and Extensions to SQL	309
Proposed Analytical Functions	310
Programming Extensions	311
Triggers and Routines	312
Triggers	312
Routines	314
Embedded SQL and Dynamic SQL	315
Summary	316
Chapter Review	317
Key Terms	317
Review Questions	317
Problems and Exercises	318
Field Exercises	319
References	319
Further Reading	319
Web Resources	319
Project Case: Mountain View Community Hospital	320

9 The Client/Server Database Environment

321

Learning Objectives	321
Introduction	321
Client/Server Architectures	322
File Server Architectures	323
Limitations of File Servers	324
Database Server Architectures	324
Three-Tier Architectures	326
Partitioning an Application	328
Role of the Mainframe	329
Using Parallel Computer Architectures	330
Multiprocessor Hardware Architectures	331
Business Related Uses of SMP and MPP Architectures	333
Using Middleware	334
Establishing Client/Server Security	336
Client/Server Security Issues for Web-Enabled Databases	336
Client/Server Issues	337

Database Access From Client Applications	339
Using Query-by-Example	340
The History and Importance of QBE	341
QBE: The Basics	341
Selecting Qualified Records	344
Self-Join	345
Basing a Query on Another Query	346
Using SQL Pass-Through Queries	348
Using ODBC to Link External Tables Stored on a Database Server	349
Using JDBC to Link External Tables Stored on a Database Server	350
Using Visual Basic for Applications (VBA) in Client Applications	354
Summary	355
Chapter Review	357
Key Terms	357
Review Questions	357
Problems and Exercises	357
Field Exercises	358
References	358
Web Resources	359
Project Case: Mountain View Community Hospital	360

10 The Internet Database Environment

363

Learning Objectives	363
Introduction	363
The Internet and Database Connection	364
The Internet Environment	365
Terminology	367
Communication-Related Terms	367
Web-Related Terms	368
Common Internet Architecture Components	368
Internet-Related Languages	368
Server-Side Extensions	370
Web Server Interfaces	370
Web Servers	371
Client-Side Extensions	372
Web-to-Database Tools: ColdFusion and ASP	373
Active Server Pages (ASP) Example	373
ColdFusion Example	377
Embedded SQL	381
Managing Website Data	383
Web Security Issues	383
Planning for Web Security	384
Network Level Security	384
Operating System Level Security	385
Web Server Security	385
Privacy Issues	385
Internet Technology Rate-of-Change Issues	387
Summary	388
Chapter Review	388

Key Terms	388
Review Questions	389
Problems and Exercises	389
Field Exercises	389
References	390
Web Resources	390

Project Case: Mountain View Community Hospital	391
---	------------

11 Data Warehousing

393

Learning Objectives	393
---------------------	-----

Introduction	393
--------------	-----

Basic Concepts of Data Warehousing	394
------------------------------------	-----

A Brief History	395
-----------------	-----

The Need for Data Warehousing	395
-------------------------------	-----

Need for a Company-Wide View	395
------------------------------	-----

Need to Separate Operational and Informational Systems	398
--	-----

Data Warehouse Architectures	399
------------------------------	-----

Generic Two-Level Architecture	399
--------------------------------	-----

Independent Data Mart Data Warehousing Environment	400
--	-----

Dependent Data Mart and Operational Data Store Architecture	401
---	-----

Logical Data Mart and @ctive Warehouse Architecture	403
---	-----

Three-Layer Data Architecture	405
-------------------------------	-----

Role of the Enterprise Data Model	405
-----------------------------------	-----

Role of Metadata	406
------------------	-----

Some Characteristics of Data Warehouse Data	406
---	-----

Status Versus Event Data	407
--------------------------	-----

Transient Versus Periodic Data	407
--------------------------------	-----

An Example of Transient and Periodic Data	408
---	-----

Transient Data	409
----------------	-----

Periodic Data	410
---------------	-----

The Reconciled Data Layer	410
---------------------------	-----

Characteristics of Data after ETL	411
-----------------------------------	-----

The ETL Process	411
-----------------	-----

Extract	412
---------	-----

Cleans	413
--------	-----

Load and Index	414
----------------	-----

Data Transformation	415
---------------------	-----

Data Transformation Functions	416
-------------------------------	-----

Record-Level Functions	416
------------------------	-----

Field-Level Functions	417
-----------------------	-----

More Complex Transformations	418
------------------------------	-----

Tools to Support Data Reconciliation	418
--------------------------------------	-----

Data Quality Tools	419
--------------------	-----

Data Conversion Tools	420
-----------------------	-----

Data-Cleansing Tools	420
----------------------	-----

The Derived Data Layer	420
------------------------	-----

Characteristics of Derived Data	420
---------------------------------	-----

The Star Schema	421
-----------------	-----

Fact Tables and Dimension Tables	421
----------------------------------	-----

Example Star Schema	421
---------------------	-----

Surrogate Key	424
---------------	-----

Grain of Fact Table	424
Size of the Fact Table	424
Modeling Date and Time	425
Variations of the Star Schema	426
Multiple Fact Tables	426
Factless Fact Tables	427
Normalizing Dimension Tables	428
Multivalued Dimensions	428
Hierarchies	429
Slowly Changing Dimensions	431
The User Interface	432
Role of Metadata	433
Querying Tools	433
On-Line Analytical Processing (OLAP) Tools	434
Slicing a Cube	435
Drill-Down	435
Data-Mining Tools	436
Data-Mining Techniques	436
Data-Mining Applications	436
Data Visualization	437
Summary	437
Chapter Review	438
Key Terms	438
Review Questions	438
Problems and Exercises	439
Field Exercises	440
References	441
Further Reading	441
Web Resources	442
Project Case: Mountain View Community Hospital	443

Part

Advanced Database Topics

PART V OVERVIEW 446

12 Data and Database Administration

Learning Objectives	447
Introduction	447
The Roles of Data and Database Administrators	448
Traditional Data Administration	448
Traditional Database Administration	449
Evolving Approaches to Data and Database Administration	449
Evolving Approaches to Data Administration	452
Data Warehouse Administration	454
Modeling Enterprise Data	455
Planning for Databases	456
Managing Data Security	457
Threats to Data Security	457
Views	459
Integrity Controls	459
Authorization Rules	461

User-Defined Procedures	462
Encryption	462
Authentication Schemes	463
Backing Up Databases	464
Basic Recovery Facilities	464
Backup Facilities	464
Journalizing Facilities	465
Checkpoint Facility	466
Recovery Manager	466
Recovery and Restart Procedures	466
Switch	466
Restore/Rerun	466
Transaction Integrity	467
Backward Recovery	468
Forward Recovery	469
Types of Database Failure	470
Aborted Transactions	470
Incorrect Data	470
System Failure	471
Database Destruction	471
Controlling Concurrent Access	471
The Problem of Lost Updates	472
Serializability	473
Locking Mechanisms	473
Locking Level	474
Types of Locks	474
Deadlock	475
Managing Deadlock	475
Versioning	476
Managing Data Quality	478
Security Policy and Disaster Recovery	478
Personnel Controls	480
Physical Access Controls	480
Maintenance Controls	480
Data Protection and Repositories	480
Data Dictionaries and Repositories	481
Repositories	481
Overview of Tuning the Database for Performance	483
Installation of the DBMS	483
Memory Usage	484
Input/Output (I/O) Contention	484
CPU Usage	485
Application Tuning	485
Summary	485
Chapter Review	486
Key Terms	486
Review Questions	486
Problems and Exercises	487
Field Exercises	489
References	490
Web Resources	490
Project Case: Mountain View Community Hospital	491

13 Distributed Databases

493

- Learning Objectives 493
- Introduction 493
 - Objectives and Trade-offs 496
- Options for Distributing a Database 498
 - Data Replication 498
 - Snapshot Replication 499
 - Near Real-Time Replication 500
 - Pull Replication 500
 - Database Integrity with Replication 501
 - When to Use Replication 501
 - Horizontal Partitioning 501
 - Vertical Partitioning 502
 - Combinations of Operations 504
 - Selecting the Right Data Distribution Strategy 505
- Distributed DBMS 506
 - Location Transparency 508
 - Replication Transparency 509
 - Failure Transparency 510
 - Commit Protocol 510
 - Concurrency Transparency 511
 - Query Optimization 512
 - Evolution of Distributed DBMS 515
 - Remote Unit of Work 515
 - Distributed Unit of Work 515
 - Distributed Request 516
- Distributed DBMS Products 516
- Summary 517
- Chapter Review 518
 - Key Terms 518
 - Review Questions 519
 - Problems and Exercises 519
 - Field Exercises 520
 - References 521
 - Web Resources 521
- Project Case: Mountain View Community Hospital 522

14 Object-Oriented Data Modeling

523

- Learning Objectives 523
- Introduction 523
- The Unified Modeling Language 526
- Object-Oriented Data Modeling 526
 - Representing Objects and Classes 526
 - Types of Operations 529
 - Representing Associations 529
 - Representing Association Classes 532
 - Representing Derived Attributes, Derived Associations, and Derived Roles 535
 - Representing Generalization 535
 - Interpreting Inheritance and Overriding 540

Representing Multiple Inheritance 541
Representing Aggregation 542

Business Rules 545

OBJECT MODELING EXAMPLE: PINE VALLEY FURNITURE COMPANY 545

Summary 548

Chapter Review 549

Key Terms 549

Review Questions 549

Problems and Exercises 550

Field Exercises 552

References 552

Web Resources 553

Project Case: Mountain View Community Hospital 554

15 Object-Oriented Database Development

556

Learning Objectives 556

Introduction 556

Object Definition Language 557

Defining a Class 557

Defining an Attribute 558

Defining User Structures 559

Defining Operations 559

Defining a Range for an Attribute 560

Defining Relationships 560

Defining an Attribute with an Object Identifier as Its Value 562

Defining Many-to-Many Relationships, Keys, and Multivalued Attributes 563

Defining Generalization 565

Defining an Abstract Class 566

Defining Other User Structures 567

OODB DESIGN FOR PINE VALLEY FURNITURE COMPANY 568

Creating Object Instances 570

Object Query Language 571

Basic Retrieval Command 571

Including Operations in Select Clause 572

Finding Distinct Values 572

Querying Multiple Classes 572

Writing Subqueries 573

Calculating Summary Values 574

Calculating Group Summary Values 574

Qualifying Groups 575

Using a Set in a Query 575

Summary of OQL 576

Current ODMS Products and Their Applications 576

Summary 577

Chapter Review 577

Key Terms 577

Review Questions 578

Problems and Exercises 578

Field Exercises 579

References 579

Web Resources 579

Project Case: Mountain View Community Hospital 580

Appendix A E-R Modeling Tools and Notation 581

- Comparing E-R Modeling Conventions 581
 - Visible Analyst Notation 581
 - Platinum ERwin (CASE Tool) Notation 582
 - Microsoft Access 2000 Notation 583
 - Oracle Designer Notation 585
- Comparison of Tool Interfaces and E-R Diagrams 586

Appendix B Advanced Normal Forms 589

- Boyce-Codd Normal Form 589
 - Anomalies in STUDENT_ADVISOR 590
 - Definition of Boyce-Codd Normal Form (BCNF) 590
 - Converting a Relation to BCNF 590
- Fourth Normal Form 591
 - Multivalued Dependencies 592
- Higher Normal Forms 593
- References 594

Appendix C Data Structures 595

- Pointers 595
- Data Structure Building Blocks 596
- Linear Data Structures 598
 - Stacks 599
 - Queues 600
 - Sorted Lists 600
 - Multilists 603
- Hazards of Chain Structures 604
- Trees 604
 - Balanced Trees 604
- Reference 607

Appendix D Object-Relational Databases 608

- Basic Concepts and Definitions 608
 - Features of an ORDBMS 609
 - Complex Data Types 609
- Enhanced SQL 609
 - A Simple Example 610
 - Content Addressing 610
- Advantages of the Object-Relational Approach 611
- ORDBMS Vendors and Products 611
- References 612
- Web Resources 612

Glossary of Acronyms	613
Glossary of Terms	615
Credits	626
Index	627