CONTENTS

Preface 23

Part I The Context of Database Management 35

An Overview of Part One 35

Chapter 1 **The Database Environment and Development Process 36**

Learning Objectives 36

Data Matter! 36

Introduction 37

Basic Concepts and Definitions 39

- Data 39
- Data Versus Information 40
- Metadata 41

Traditional File Processing Systems 42

- File Processing Systems at Pine Valley Furniture Company 42
- Disadvantages of File Processing Systems 42
  - Program-Data Dependence 43
  - Duplication of Data 43
  - Limited Data Sharing 43
  - Lengthy Development Times 43
  - Excessive Program Maintenance 43

The Database Approach 44

- Data Models 44
  - Entities 44
  - Relationships 44
- Relational Databases 44
- Database Management Systems 45
- Advantages of the Database Approach 46
  - Program-Data Independence 47
  - Planned Data Redundancy 47
  - Improved Data Consistency 47
  - Improved Data Sharing 47
  - Increased Productivity of Application Development 47
  - Enforcement of Standards 47
  - Improved Data Quality 48
  - Improved Data Accessibility and Responsiveness 48
  - Reduced Program Maintenance 48
  - Improved Decision Support 49
- Cautions About Database Benefits 49
- Costs and Risks of the Database Approach 49
  - New, Specialized Personnel 49
  - Installation and Management Cost and Complexity 49
  - Conversion Costs 49
  - Need for Explicit Backup and Recovery 49
  - Organizational Conflict 50

Components of the Database Environment 50

The Database Development Process 51
Systems Development Life Cycle 52
PLANNING—ENTERPRISE MODELING 53
PLANNING—CONCEPTUAL DATA MODELING 53
ANALYSIS—CONCEPTUAL DATA MODELING 54
DESIGN—LOGICAL DATABASE DESIGN 54
DESIGN—PHYSICAL DATABASE DESIGN AND DEFINITION 54
IMPLEMENTATION—DATABASE IMPLEMENTATION 55
MAINTENANCE—DATABASE MAINTENANCE 55
Alternative Information Systems (IS) Development Approaches 55
Three-Schema Architecture for Database Development 57
Managing the People Involved in Database Development 57
Evolution of Database Systems 59
1960s 59
1970s 61
1980s 61
1990s 61
2000 and Beyond 61
The Range of Database Applications 62
Personal Databases 62
Two-Tier Client/Server Databases 63
Multitier Client/Server Databases 63
Enterprise Applications 64
Developing a Database Application for Pine Valley Furniture Company 65
Database Evolution at Pine Valley Furniture Company 67
Project Planning 67
Analyzing Database Requirements 68
Designing the Database 70
Using the Database 73
Administering the Database 75
Future of Databases at Pine Valley 75
Summary 75 • Key Terms 76 • Review Questions 77 • Problems and Exercises 78 • Field Exercises 79 • References 80 • Further Reading 80 • Web Resources 81
▶ **CASE: Mountain View Community Hospital 82**

Part II Database Analysis 87
An Overview of Part Two 87

Chapter 2 **Modeling Data in the Organization 89**
Learning Objectives 89
Introduction 89
The E-R Model: An Overview 91
Sample E-R Diagram 91
E-R Model Notation 93
Modeling the Rules of the Organization 94
Overview of Business Rules 95
THE BUSINESS RULES PARADIGM 95

Scope of Business Rules 96
Good Business Rules 96
Gathering Business Rules 97
Data Names and Definitions 97
Data Names 97
Data Definitions 98
Good Data Definitions 98
Modeling Entities and Attributes 100
Entities 100
Entity Type Versus Entity Instance 100
Entity Type Versus System Input, Output, or User 100
Strong Versus Weak Entity Types 101
Naming and Defining Entity Types 102
Attributes 104
Required Versus Optional Attributes 104
Simple Versus Composite Attributes 105
Single-Valued Versus Multivalued Attributes 105
Stored Versus Derived Attributes 106
Identifier Attribute 106
Naming and Defining Attributes 107
Modeling Relationships 109
Basic Concepts and Definitions in Relationships 110
Attributes on Relationships 111
Associative Entities 111
Degree of a Relationship 113
Unary Relationship 113
Binary Relationship 115
Ternary Relationship 116
Attributes or Entity? 117
Cardinality Constraints 119
Minimum Cardinality 119
Maximum Cardinality 119
Some Examples of Relationships and Their Cardinalities 120
A Ternary Relationship 121
Modeling Time-Dependent Data 121
Modeling Multiple Relationships Between Entity Types 124
Naming and Defining Relationships 126
E-R Modeling Example: Pine Valley Furniture Company 127
Database Processing at Pine Valley Furniture 130
Showing Product Information 130
Showing Product Line Information 131
Showing Customer Order Status 131
Showing Product Sales 132
Summary 133 • Key Terms 134 • Review Questions 134 • Problems and Exercises 135 • Field Exercises 142 • References 143 • Further Reading 143 • Web Resources 143
▶ **CASE: Mountain View Community Hospital 144**

Chapter 3 The Enhanced E-R Model 146
Learning Objectives 146
Introduction 146

Representing Supertypes and Subtypes 147
Basic Concepts and Notation 148
An Example of a Supertype/Subtype Relationship 149
Attribute Inheritance 150
When to Use Supertype/Subtype Relationships 150
Representing Specialization and Generalization 151
Generalization 151
Specialization 153
Combining Specialization and Generalization 154
Specifying Constraints in Supertype/Subtype Relationships 154
Specifying Completeness Constraints 154
Total Specialization Rule 154
Partial Specialization Rule 155
Specifying Disjointness Constraints 155
Disjoint Rule 155
Overlap Rule 156
Defining Subtype Discriminators 156
Disjoint Subtypes 157
Overlapping Subtypes 157
Defining Supertype/Subtype Hierarchies 158
An Example of a Supertype/Subtype Hierarchy 158
Summary of Supertype/Subtype Hierarchies 159
EER Modeling Example: Pine Valley Furniture Company 160
Entity Clustering 163
Packaged Data Models 166
A Revised Data Modeling Process with Packaged Data Models 168
Packaged Data Model Examples 170
Summary 175 • Key Terms 176 • Review Questions 176 • Problems and Exercises 177 • Field Exercises 180 • References 180 • Further Reading 180 • Web Resources 181
► **CASE: Mountain View Community Hospital 182**

Part III Database Design 185
An Overview of Part Three 185

Chapter 4 **Logical Database Design and the Relational Model 187**
Learning Objectives 187
Introduction 187
The Relational Data Model 188
Basic Definitions 188
Relational Data Structure 189
Relational Keys 189
Properties of Relations 190
Removing Multivalued Attributes from Tables 190
Sample Database 190
Integrity Constraints 192
Domain Constraints 192
Entity Integrity 192
Referential Integrity 194

Creating Relational Tables 195
Well-Structured Relations 196
Transforming EER Diagrams into Relations 197
Step 1: Map Regular Entities 198
Composite Attributes 198
Multivalued Attributes 199
Step 2: Map Weak Entities 199
When to Create a Surrogate Key 200
Step 3: Map Binary Relationships 201
Map Binary One-to-Many Relationships 201
Map Binary Many-to-Many Relationships 201
Map Binary One-to-One Relationships 202
Step 4: Map Associative Entities 203
Identifier Not Assigned 203
Identifier Assigned 204
Step 5: Map Unary Relationships 205
Unary One-to-Many Relationships 205
Unary Many-to-Many Relationships 205
Step 6: Map Ternary (and n-ary) Relationships 207
Step 7: Map Supertype/Subtype Relationships 207
Summary of EER-to-Relational Transformations 209
Introduction to Normalization 210
Steps in Normalization 211
Functional Dependencies and Keys 212
Determinants 212
Candidate Keys 212
Normalization Example: Pine Valley Furniture Company 214
Step 0: Represent the View in Tabular Form 214
Step 1: Convert to First Normal Form 215
Remove Repeating Groups 215
Select the Primary Key 215
Anomalies in 1NF 216
Step 2: Convert to Second Normal Form 216
Step 3: Convert to Third Normal Form 217
Removing Transitive Dependencies 218
Determinants and Normalization 219
Step 4: Further Normalization 219
Merging Relations 219
An Example 220
View Integration Problems 220
Synonyms 220
Homonyms 221
Transitive Dependencies 221
Supertype/Subtype Relationships 222
A Final Step for Defining Relational Keys 222
Summary 224 • Key Terms 226 • Review Questions 226 • Problems and Exercises 227 • Field Exercises 234 • References 234 • Further Reading 235 • Web Resources 235
▶ **CASE: Mountain View Community Hospital 236**

Chapter 5 **Physical Database Design and Performance 240**
Learning Objectives 240
Introduction 240
The Physical Database Design Process 241
Physical Database Design as a Basis for Regulatory Compliance 242
Data Volume and Usage Analysis 243
Designing Fields 244
Choosing Data Types 245
Coding Techniques 246
Handling Missing Data 247
Denormalizing and Partitioning Data 247
Denormalization 247
Opportunities for and Types of Denormalization 248
Denormalize With Caution 250
Partitioning 251
Designing Physical Database Files 253
File Organizations 254
Sequential File Organizations 255
Indexed File Organizations 255
Hashed File Organizations 258
Clustering Files 260
Designing Controls for Files 261
Using and Selecting Indexes 262
Creating a Unique Key Index 262
Creating a Secondary (Nonunique) Key Index 262
When to Use Indexes 263
Designing a Database for Optimal Query Performance 264
Parallel Query Processing 264
Overriding Automatic Query Optimization 265
Summary 266 • Key Terms 267 • Review Questions 267 • Problems and Exercises 267 • Field Exercises 270 • References 270 • Further Reading 271 • Web Resources 271

▶ **CASE: Mountain View Community Hospital 272**

Part IV Implementation 275
An Overview of Part Four 275

Chapter 6 **Introduction to SQL 277**
Learning Objectives 277
Introduction 277
Origins of the SQL Standard 279
The SQL Environment 280
Defining a Database in SQL 285
Generating SQL Database Definitions 286
Creating Tables 286
Creating Data Integrity Controls 289
Changing Table Definitions 290
Removing Tables 291

THE CENTER FOR LIBRARY RESOURCES AND EDUCATIONAL MEDIA
SURANAREE UNIVERSITY OF TECHNOLOGY

Inserting, Updating, and Deleting Data 291
Batch Input 292
Deleting Database Contents 293
Updating Database Contents 293
Internal Schema Definition in RDBMSs 294
Creating Indexes 294
Processing Single Tables 295
Clauses of the SELECT Statement 295
Using Expressions 297
Using Functions 298
Using Wildcards 301
Using Comparison Operators 301
Using Null Values 302
Using Boolean Operators 302
Using Ranges for Qualification 304
Using Distinct Values 305
Using IN and NOT IN with Lists 307
Sorting Results: The ORDER BY Clause 307
Categorizing Results: The GROUP BY Clause 308
Qualifying Results by Categories: The HAVING Clause 309
Using and Defining Views 311
MATERIALIZED VIEWS 315
Summary 315 • Key Terms 316 • Review Questions 316 • Problems and Exercises 317 • Field Exercises 320 • References 320 • Further Reading 321 • Web Resources 321
▶ **CASE: Mountain View Community Hospital 322**

Chapter 7 **Advanced SQL 323**

Learning Objectives 323
Introduction 323
Processing Multiple Tables 324
Equi-join 325
Natural Join 326
Outer Join 327
Sample Join Involving Four Tables 329
Self-Join 331
Subqueries 332
Correlated Subqueries 337
Using Derived Tables 339
Combining Queries 340
Conditional Expressions 342
More Complicated SQL Queries 342
Tips for Developing Queries 344
Guidelines for Better Query Design 346
Ensuring Transaction Integrity 348
Data Dictionary Facilities 349
SQL:2008 Enhancements and Extensions to SQL 351

Analytical and OLAP Functions 351
New Data Types 353
Other Enhancements 353
Programming Extensions 354
Triggers and Routines 355
Triggers 356
Routines 358
Example Routine in Oracle's PL/SQL 359
Embedded SQL and Dynamic SQL 361
Summary 363 • Key Terms 364 • Review Questions 364 • Problems and Exercises 365 • Field Exercises 368 • References 368 • Further Reading 368 • Web Resources 368

▶ CASE: Mountain View Community Hospital 369

Chapter 8 Database Application Development 370
Learning Objectives 370
Location, Location, Location! 370
Introduction 371
Client/Server Architectures 371
Databases in a Two-Tier Architecture 373
A VB.NET Example 375
A Java Example 377
Three-Tier Architectures 378
Web Application Components 380
Databases in Three-Tier Applications 382
A JSP Web Application 382
A PHP Example 386
An ASP.NET Example 386
Key Considerations in Three-Tier Applications 388
Stored Procedures 389
Transactions 390
Database Connections 392
Key Benefits of Three-Tier Applications 392
Cloud Computing and Three-Tier Applications 393
Extensible Markup Language (XML) 394
Storing XML Documents 396
Retrieving XML Documents 396
Displaying XML Data 399
XML and Web Services 399
Summary 403 • Key Terms 403 • Review Questions 404 • Problems and Exercises 404 • Field Exercises 405 • References 405 • Further Reading 405 • Web Resources 405

▶ CASE: Mountain View Community Hospital 407

Chapter 9 Data Warehousing 409
Learning Objectives 409
Introduction 409
Basic Concepts of Data Warehousing 411

A Brief History of Data Warehousing 412
The Need for Data Warehousing 412
Need for a Company-Wide View 412
Need to Separate Operational and Informational Systems 415
Data Warehouse Architectures 415
Independent Data Mart Data Warehousing Environment 415
Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach 417
Logical Data Mart and Real-Time Data Warehouse Architecture 419
Three-Layer Data Architecture 422
Role of the Enterprise Data Model 423
Role of Metadata 423
Some Characteristics of Data Warehouse Data 423
Status Versus Event Data 423
Transient Versus Periodic Data 424
An Example of Transient and Periodic Data 424
Transient Data 424
Periodic Data 426
Other Data Warehouse Changes 426
The Derived Data Layer 427
Characteristics of Derived Data 427
The Star Schema 428
Fact Tables and Dimension Tables 428
Example Star Schema 429
Surrogate Key 431
Grain of the Fact Table 431
Duration of the Database 432
Size of the Fact Table 432
Modeling Date and Time 433
Variations of the Star Schema 434
Multiple Fact Tables 434
Factless Fact Tables 435
Normalizing Dimension Tables 436
Multivalued Dimensions 436
Hierarchies 437
Slowly Changing Dimensions 439
Determining Dimensions and Facts 441
Big Data and Columnar Databases 443
If You Knew SQL Like I NoSQL 445
The User Interface 445
Role of Metadata 446
SQL OLAP Querying 446
Online Analytical Processing (OLAP) Tools 448
Slicing a Cube 448
Drill-Down 449
Summarizing More than Three Dimensions 450
Data Visualization 450
Business Performance Management and Dashboards 452
Data-Mining Tools 453

Data-Mining Techniques 453
Data-Mining Applications 453
Summary 454 • Key Terms 455 • Review Questions 455 • Problems and Exercises 456 • Field Exercises 460 • References 460 • Further Reading 461 • Web Resources 461

► **CASE: Mountain View Community Hospital 463**

Part V Advanced Database Topics 467
An Overview of Part Five 467

Chapter 10 **Data Quality and Integration 469**
Learning Objectives 469
Introduction 469
Data Governance 470
Managing Data Quality 471
Characteristics of Quality Data 472
External Data Sources 473
Redundant Data Storage and Inconsistent Metadata 474
Data Entry Problems 474
Lack of Organizational Commitment 474
Data Quality Improvement 474
Get the Business Buy-In 474
Conduct a Data Quality Audit 475
Establish a Data Stewardship Program 476
Improve Data Capture Processes 476
Apply Modern Data Management Principles and Technology 477
Apply TQM Principles and Practices 477
Summary of Data Quality 477
Master Data Management 478
Data Integration: An Overview 479
General Approaches to Data Integration 479
Data Federation 480
Data Propagation 481
Data Integration for Data Warehousing: The Reconciled Data Layer 481
Characteristics of Data After ETL 481
The ETL Process 482
Mapping and Metadata Management 482
Extract 483
Cleanse 484
Load and Index 486
Data Transformation 487
Data Transformation Functions 488
Record-Level Functions 488
Field-Level Functions 489
Summary 491 • Key Terms 491 • Review Questions 491 • Problems and Exercises 492 • Field Exercises 493 • References 493 • Further Reading 494 • Web Resources 494

► **CASE: Mountain View Community Hospital 495**

Chapter 11 **Data and Database Administration 497**
Learning Objectives 497
Introduction 497
The Roles of Data and Database Administrators 498
Traditional Data Administration 499
Traditional Database Administration 500
Trends in Database Administration 503
Data Warehouse Administration 504
Summary of Evolving Data Administration Roles 504
The Open Source Movement and Database Management 505
Managing Data Security 506
Threats to Data Security 507
Establishing Client/Server Security 508
Server Security 508
Network Security 509
Application Security Issues in Three-Tier Client/Server Environments 509
Data Privacy 510
Database Software Data Security Features 511
Views 512
Integrity Controls 513
Authorization Rules 514
User-Defined Procedures 516
Encryption 516
Authentication Schemes 517
Passwords 517
Strong Authentication 517
Sarbanes-Oxley (SOX) and Databases 518
IT Change Management 518
Logical Access to Data 518
Personnel Controls 519
Physical Access Controls 519
IT Operations 519
Database Backup and Recovery 520
Basic Recovery Facilities 520
Backup Facilities 520
Journalizing Facilities 521
Checkpoint Facility 521
Recovery Manager 522
Recovery and Restart Procedures 522
Disk Mirroring 522
Restore/Rerun 522
Maintaining Transaction Integrity 522
Backward Recovery 524
Forward Recovery 524
Types of Database Failure 525
Aborted Transactions 526

Incorrect Data 526
System Failure 526
Database Destruction 526
Disaster Recovery 527
Controlling Concurrent Access 527
The Problem of Lost Updates 527
Serializability 528
Locking Mechanisms 528
Locking Level 529
Types of Locks 530
Deadlock 530
Managing Deadlock 531
Versioning 532
Data Dictionaries and Repositories 533
Data Dictionary 533
Repositories 533
Overview of Tuning the Database for Performance 535
Installation of the DBMS 535
Memory and Storage Space Usage 536
Input/Output (I/O) Contention 536
CPU Usage 537
Application Tuning 537
Data Availability 538
Costs of Downtime 538
Measures to Ensure Availability 539
Hardware Failures 539
Loss or Corruption of Data 539
Human Error 539
Maintenance Downtime 539
Network-Related Problems 540
Summary 540 • Key Terms 541 • Review Questions 541 • Problems and Exercises 542 • Field Exercises 544 • References 545 • Further Reading 545 • Web Resources 545

▶ **CASE: Mountain View Community Hospital 546**

Chapter 12 Overview: Distributed Databases 547
Learning Objectives 547
Overview 547
Objectives and Trade-offs 548
Options for Distributing a Database 548
Distributed DBMS 549
Query Optimization 549
Summary 550 • Chapter Review 551 • References 551 • Further Reading 551 • Web Resources 551

Chapter 13 Overview: Object-Oriented Data Modeling 552
Learning Objectives 552
Overview 552
Unified Modeling Language 553
Object-Oriented Data Modeling 553

Representing Aggregation 559
Summary 559 • Chapter Review 560 • References 560 • Further Reading 561 • Web Resources 561

Chapter 14 **Overview: Using Relational Databases to Provide Object Persistence 562**
Learning Objectives 562
Overview 562
Providing Persistence for Objects Using Relational Databases 563
Call-Level Application Programming Interfaces 563
SQL Query Mapping Frameworks 564
Object-Relational Mapping Frameworks 564
Proprietary Approaches 564
Selecting the Right Approach 564
Object-Relational Mapping Example 565
Mapping Files 566
Responsibilities of Object-Relational Mapping Frameworks 569
Summary 570 • Chapter Review 570 • References 570 • Further Reading 571 • Web Resources 571

Appendix A **Data Modeling Tools and Notation 572**
Comparing E-R Modeling Conventions 572
Visio Professional 2010 Notation 572
Entities 576
Relationships 576
CA ERwin Data Modeler r8 Notation 576
Entities 576
Relationships 576
Sybase PowerDesigner 50 Notation 578
Entities 579
Relationships 579
Oracle Designer Notation 579
Entities 579
Relationships 579
Comparison of Tool Interfaces and E-R Diagrams 579

Appendix B **Advanced Normal Forms 582**
Boyce-Codd Normal Form 582
Anomalies in Student Advisor 582
Definition of Boyce-Codd Normal Form (BCNF) 583
Converting a Relation to BCNF 583
Fourth Normal Form 584
Multivalued Dependencies 586
Higher Normal Forms 586
Key Terms 587 • References 587 • Web Resources 587

Appendix C **Data Structures 588**
Pointers 588
Data Structure Building Blocks 589
Linear Data Structures 591
Stacks 592

Queues 592
Sorted Lists 593
Multilists 595
Hazards of Chain Structures 595
Trees 596
Balanced Trees 596
Reference 599

Glossary of Acronyms 600
Glossary of Terms 602
Index 610

ONLINE CHAPTERS

Chapter 12 **Distributed Databases 12-1**
Learning Objectives 12-1
Introduction 12-1
Objectives and Trade-offs 12-4
Options for Distributing a Database 12-6
Data Replication 12-6
Snapshot Replication 12-7
Near-Real-Time Replication 12-8
Pull Replication 12-8
Database Integrity With Replication 12-8
When to Use Replication 12-8
Horizontal Partitioning 12-9
Vertical Partitioning 12-10
Combinations of Operations 12-11
Selecting the Right Data Distribution Strategy 12-11
Distributed DBMS 12-13
Location Transparency 12-15
Replication Transparency 12-16
Failure Transparency 12-17
Commit Protocol 12-17
Concurrency Transparency 12-18
Time-Stamping 12-18
Query Optimization 12-19
Evolution of Distributed DBMSs 12-21
Remote Unit of Work 12-22
Distributed Unit of Work 12-22
Distributed Request 12-23
Distributed DBMS Products 12-23
Summary 12-24 • Key Terms 12-25 • Review Questions 12-25 • Problems and Exercises 12-26 • Field Exercises 12-27 • References 12-28 • Further Reading 12-28 • Web Resources 12-28

Chapter 13 **Object-Oriented Data Modeling 13-1**
Learning Objectives 13-1
Introduction 13-1
Unified Modeling Language 13-3
Object-Oriented Data Modeling 13-4
Representing Objects and Classes 13-4
Types of Operations 13-6
Representing Associations 13-7
Representing Association Classes 13-10
Representing Derived Attributes, Derived Associations, and Derived Roles 13-12
Representing Generalization 13-12
Interpreting Inheritance and Overriding 13-17
Representing Multiple Inheritance 13-18
Representing Aggregation 13-19

Business Rules 13-22

Object Modeling Example: Pine Valley Furniture Company 13-23

Summary 13-25 • Key Terms 13-26 • Review Questions 13-26 • Problems and Exercises 13-29 • Field Exercises 13-35 • References 13-35 • Further Reading 13-36 • Web Resources 13-36

Chapter 14 Using Relational Databases to Provide Object Persistence 14-1

Learning Objectives 14-1

Introduction 14-1

Object-Relational Impedance Mismatch 14-3

Providing Persistence for Objects Using Relational Databases 14-6

Common Approaches 14-6

Call-Level Application Programming Interfaces 14-6

SQL Query Mapping Frameworks 14-6

Object-Relational Mapping Frameworks 14-7

Proprietary Approaches 14-7

Selecting the Right Approach 14-7

Call-Level APIs 14-8

SQL Query Mapping Frameworks 14-9

ORM Frameworks 14-9

Object-Relational Mapping Example Using Hibernate 14-10

Foundation 14-10

Mapping Files 14-12

Hibernate Configuration 14-14

Mapping Object-Oriented Structures to a Relational Database 14-16

Class 14-16

Inheritance: Superclass–Subclass 14-16

One-to-One Association 14-18

Many-to-One and One-to-Many Associations 14-18

Aggregation and Composition 14-18

Many-to-Many Associations 14-19

Responsibilities of Object-Relational Mapping Frameworks 14-19

HQL 14-20

Summary 14-24 • Key Terms 14-25 • Review Questions 14-25 • Problems and Exercises 14-26 • Field Exercises 14-26 • References 14-26 • Further Reading 14-27 • Web Resources 14-27