Modern Database Management
Fifth Edition

FRED R. McFADDEN
University of Colorado–Colorado Springs

JEFFREY A. HOFFER
University of Dayton

MARY B. PRESCOTT
University of South Florida

An imprint of Addison Wesley Longman, Inc.
Contents

Preface xix

Part I The Context of Database Management 1

PART I OVERVIEW 2

Chapter 1 The Database Environment 3

Learning Objectives 3

Introduction 3

Basic Concepts and Definitions 4

Data 4

Data versus Information 5

Metadata 5

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 9

Limited Data Sharing 9

Lengthy Development Times 9

Excessive Program Maintenance 10

The Database Approach 10

The Database Approach at Pine Valley Furniture Company 10

Enterprise Data Model 10

Relational Databases 11

Implementing the Relational Databases 13

A Database Application 14

The Range of Database Applications 15

Personal Computer Databases 15

Workgroup Databases 16

Department Databases 18

Enterprise Databases 19

Summary of Database Applications 20

Advantages of the Database Approach 20

Program-Data Independence 21

Minimal Data Redundancy 21
Moved to the second page
Alternative IS Development Approaches  48
The Role of CASE and a Repository  50
Managing the People Involved in Database Development  51
Three-Schema Architecture for Database Development  53
Three-Tiered Database Location Architecture  56
Developing a Database Application for Pine Valley Furniture  58
Matching User Needs to the Information Systems Architecture  59
Analyzing Database Requirements  61
Designing the Database  64
Using a Database  67
Administering a Database  69
Summary  70
Chapter Review  71
Key Terms  71
Review Questions  71
Problems and Exercises  72
Field Exercises  75
References  75
Further Reading  76
Project Case: Mountain View Community Hospital  77

Part II Database Analysis  83

Chapter 3 The Entity-Relationship Model  85
Learning Objectives  85
Introduction  85
The E-R Model  87
Sample E-R Diagram  87
E-R Model Notation  89
Entity-Relationship Model Constructs  89
Entities  89
Entity Type Versus Entity Instance  91
Entity Type Versus System Input, Output, or User  91
Strong Versus Weak Entity Types  92
Attributes  93
Simple Versus Composite Attributes  94
Single-Valued Versus Multivalued Attributes  95
Stored Versus Derived Attributes  95
Relationships  97
Basic Concepts and Definitions in Relationships  98
Attributes on Relationships  99
Associative Entities  99
Degree of a Relationship  101
Unary Relationship  101
Business Rules: An Overview 145
The Business Rules Paradigm 146
Scope of Business Rules 146
Classification of Business Rules 147
Business Rules: Defining Structural Constraints 148
Definitions 148
Facts 148
Derived Facts 149
Definitions for Data Model 150
Importance of Precise Definitions 150
Domain Constraints 153
Business Rules: Defining Operational Constraints 154
Declarative Approach to Business Rules 154
Constraint Specification Language 155
Constrained Objects and Constraining Objects 155
Sample Business Rules 156
Summary 159
Chapter Review 160
Key Terms 160
Review Questions 160
Problems and Exercises 161
Field Exercises 163
References 164
Further Reading 164

Project Case: Mountain View Community Hospital 165

Chapter 5
Object-Oriented Modeling 167
Learning Objectives 167
Introduction 167
The Unified Modeling Language 170
Object-Oriented Modeling 171
Representing Objects and Classes 171
Types of Operations 173
Representing Associations 174
Representing Association Classes 177
Representing Derived Attributes, Derived Associations, and Derived Roles 180
Representing Generalization 181
Interpreting Inheritance and Overriding 186
Representing Multiple Inheritance 187
Representing Aggregation 187
Business Rules 191
Object Modeling Example: Pine Valley Furniture Company 191
Summary 194
Chapter Review 195
Key Terms 195
Review Questions 196
Problems and Exercises 197
Part III Database Design 205

PART III OVERVIEW 206

Chapter 6 Logical Database Design and the Relational Model 207

Learning Objectives 207

Introduction 207

The Relational Data Model 208

Basic Definitions 208

Relational Data Structure 209

Relational Keys 209

Properties of Relations 210

Removing Multivalued Attributes from Tables 210

Example Database 211

Integrity Constraints 213

Domain Constraints 213

Entity Integrity 213

Referential Integrity 214

Operational Constraints 215

Creating Relational Tables 215

Well-Structured Relations 217

Transforming EER Diagrams into Relations 218

Step 1: Map Regular Entities 219

Composite Attributes 219

Multivalued Attributes 220

Step 2: Map Weak Entities 221

Step 3: Map Binary Relationships 222

Map Binary One-to-Many Relationships 222

Map Binary Many-to-Many Relationships 223

Map Binary One-to-One Relationships 224

Step 4: Map Associative Entities 224

Identifier Not Assigned 224

Identifier Assigned 225

Step 5: Map Unary Relationships 227

Unary One-to-Many Relationships 227

Unary Many-to-Many Relationships 228

Step 6: Map Ternary (and n-ary) Relationships 229

Step 7: Map Supertype/Subtype Relationships 231

Introduction to Normalization 232

Steps in Normalization 233

Functional Dependencies and Keys 235

Determinants 235

Candidate Keys 235

The Basic Normal Forms 237

First Normal Form 237

Second Normal Form 237
Chapter 7

**Physical Database Design** 253

**Learning Objectives** 253

**Introduction** 253

**Physical Database Design Process** 254

- **Data Volume and Usage Analysis** 255

**Designing Fields** 257

- **Choosing Data Types** 257
  - **Coding and Compression Techniques** 257
  - **Controlling Data Integrity** 259
  - **Handling Missing Data** 260

**Designing Physical Records and Denormalization** 260

- **Handling Fixed-Length Fields** 261
- **Handling Variable-Length Fields** 261
- **Denormalization** 261

**Designing Physical Files** 267

- **Pointer** 267
- **Access Methods** 267
- **File Organizations** 268
  - **Sequential File Organizations** 268
  - **Indexed File Organizations** 268
  - **Hashed File Organizations** 272
- **Summary of File Organizations** 272
- **Clustering Files** 272
- **Designing Controls for Files** 274

**Using and Selecting Indexes** 274

- **Creating a Primary Key Index** 274
- **Creating a Secondary Key Index** 275
- **When to Use Indexes** 275

**RAID: Improving File Access Performance by Parallel Processing** 276
Choosing Among RAID Levels 278
  RAID-0 278
  RAID-1 278
  RAID-2 281
  RAID-3 281
  RAID-4 281
  RAID-5 281
  RAID Performance 281
Designing Databases 282
  Choosing Database Architectures 282
Optimizing for Query Performance 285
Summary 286
Chapter Review 288
  Key Terms 288
  Review Questions 288
  Problems and Exercises 289
  Field Exercises 291
  References 292
  Further Reading 292
Project Case: Mountain View Community Hospital 293

Part IV Implementation 295

Part IV Overview 296

Chapter 8 Client/Server and Middleware 297
Learning Objectives 297
Introduction 297
Client/Server Architectures 298
  File Server Architectures 299
  Limitations of File Servers 300
  Database Server Architectures 301
Three-Tier Architectures 302
Partitioning an Application 304
Role of the Mainframe 305
Using Parallel Computer Architectures 306
  Multiprocessor Hardware Architectures 307
  Business Related Uses of SMP and MPP Architectures 308
Using Middleware 309
Establishing Client/Server Security 311
Client/Server Issues 312
Summary 314
Chapter Review 315
  Key Terms 315
  Review Questions 316
  Problems and Exercises 316
Chapter 9

SQL 323

Learning Objectives 323

Introduction 323

History of the SQL Standard 324

The Role of SQL in a Database Architecture 325

The SQL Environment 327

Defining a Database in SQL 331
  Generating SQL Database Definitions 332
  Creating Tables 332
  Creating Data Integrity Controls 334
  Changing Table Definitions 336
  Removing Tables 336
  Establishing Synonyms 336

Inserting, Updating, and Deleting Data 337
  Batch Input 337
  Deleting Database Contents 338
  Changing Database Contents 338

Internal Schema Definition in RDBMSs 338
  Creating Indexes 339

Processing Single Tables 340
  Clauses of the SELECT Statement 340
  Using Expressions 342
  Using Functions 343
  Using Wildcards 345
  Comparison Operators 345
  Using Boolean Operators 346
  Ranges 347
  Distinct 348
  IN and NOT IN Lists 350
  Sorting Results: The ORDER BY Clause 350
  Categorizing Results: The GROUP BY Clause 351
  Qualifying Results by Categories: The HAVING Clause 352

Processing Multiple Tables 354
  Equi-join 354
  Natural Join 355
  Outer Join 356
  Subqueries 358
  Correlated Subqueries 361

View Definitions 363

Ensuring Transaction Integrity 366

Data Dictionary Facilities 368

Triggers and Procedures 369

SQL3 371
Chapter 10  
**Database Access from Client Applications**  381

Learning Objectives  381

Introduction  381

Survey of Desktop Database Technology  382

Using Query-by-Example  383

- The History and Importance of QBE  384
- QBE: The Basics  384
- Database Definition  386
- Relationships  387
- Building Queries Using QBE  388
- Single-Table Queries  388
- Selecting Qualified Records  390
- Multiple-Table Queries  390
- Self-Join  392
- Basing a Query on Another Query  393
- Access97 Query Types  394

Building a Client Application  396

- Application Menus  396
- Form Development  397
- Report Development  399

Using OLE, COM, and ActiveX Controls for Database Access  401

Embedding SQL in Programs  403

Using Visual Basic for Applications (VBA) in Client Applications  406

Building Internet Database Servers  407

Summary  409

Chapter Review  410

- Key Terms  410
- Review Questions  410
- Problems and Exercises  411
- Field Exercises  413
- References  414

Project Case: Mountain View Community Hospital  415

Chapter 11  
**Distributed Databases**  417

Learning Objectives  417
<table>
<thead>
<tr>
<th>Introduction</th>
<th>417</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives and Trade-offs</td>
<td>421</td>
</tr>
<tr>
<td>Options for Distributing a Database</td>
<td>422</td>
</tr>
<tr>
<td>Data Replication</td>
<td>423</td>
</tr>
<tr>
<td>Snapshot Replication</td>
<td>424</td>
</tr>
<tr>
<td>Near Real-Time Replication</td>
<td>425</td>
</tr>
<tr>
<td>Pull Replication</td>
<td>425</td>
</tr>
<tr>
<td>Database Integrity with Replication</td>
<td>425</td>
</tr>
<tr>
<td>When to Use Replication</td>
<td>426</td>
</tr>
<tr>
<td>Horizontal Partitioning</td>
<td>426</td>
</tr>
<tr>
<td>Vertical Partitioning</td>
<td>427</td>
</tr>
<tr>
<td>Combinations of Operations</td>
<td>429</td>
</tr>
<tr>
<td>Selecting the Right Data Distribution Strategy</td>
<td>430</td>
</tr>
<tr>
<td>Distributed DBMS</td>
<td>431</td>
</tr>
<tr>
<td>Location Transparency</td>
<td>433</td>
</tr>
<tr>
<td>Replication Transparency</td>
<td>434</td>
</tr>
<tr>
<td>Failure Transparency</td>
<td>435</td>
</tr>
<tr>
<td>Commit Protocol</td>
<td>435</td>
</tr>
<tr>
<td>Concurrency Transparency</td>
<td>436</td>
</tr>
<tr>
<td>Timestamping</td>
<td>437</td>
</tr>
<tr>
<td>Query Optimization</td>
<td>437</td>
</tr>
<tr>
<td>Evolution of Distributed DBMS</td>
<td>440</td>
</tr>
<tr>
<td>Remote Unit of Work</td>
<td>440</td>
</tr>
<tr>
<td>Distributed Unit of Work</td>
<td>441</td>
</tr>
<tr>
<td>Distributed Request</td>
<td>441</td>
</tr>
<tr>
<td>Distributed DBMS Products</td>
<td>442</td>
</tr>
<tr>
<td>IBM Corporation and DB2</td>
<td>442</td>
</tr>
<tr>
<td>Sybase Inc.</td>
<td>443</td>
</tr>
<tr>
<td>Oracle Corporation</td>
<td>444</td>
</tr>
<tr>
<td>Computer Associates International and Ingres</td>
<td>444</td>
</tr>
<tr>
<td>Microsoft Corporation and SQL Server</td>
<td>445</td>
</tr>
<tr>
<td>Summary</td>
<td>445</td>
</tr>
<tr>
<td>Chapter Review</td>
<td>447</td>
</tr>
<tr>
<td>Key Terms</td>
<td>447</td>
</tr>
<tr>
<td>Review Questions</td>
<td>447</td>
</tr>
<tr>
<td>Problems and Exercises</td>
<td>448</td>
</tr>
<tr>
<td>Field Exercises</td>
<td>450</td>
</tr>
<tr>
<td>References</td>
<td>451</td>
</tr>
<tr>
<td>Project Case: Mountain View Community Hospital</td>
<td>452</td>
</tr>
</tbody>
</table>

Chapter 12

Object-Oriented Database Development | 453

Learning Objectives | 453

Introduction | 453

Object Definition Language | 454

Defining a Class | 454
Defining an Attribute | 455
Defining User Structures | 456
Defining Operations | 456
Defining a Range for an Attribute | 457
Defining Relationships | 457
Defining an Attribute with an Object Identifier as Its Value | 459
Defining Many-to-Many Relationships, Keys, and Multivalued Attributes 460
Defining Generalization 462
Defining an Abstract Class 464
Defining Other User Structures 464
OODB Design for Pine Valley Furniture Company 466
Creating Object Instances 467
Object Query Language 468
  Basic Retrieval Command 469
  Including Operations in Select Clause 469
  Finding Distinct Values 470
  Querying Multiple Classes 470
  Writing Subqueries 471
  Calculating Summary Values 471
  Calculating Group Summary Values 472
  Qualifying Groups 472
  Using a Set in a Query 473
Current ODBMS Products and Their Applications 474
Summary 474
Chapter Review 475
  Key Terms 475
  Review Questions 476
  Problems and Exercises 476
  Field Exercises 477
  References 478
Project Case: Mountain View Community Hospital 479

Part V Data Administration 481

Part V Overview 482

Chapter 13 Data and Database Administration 483
Learning Objectives 483
Introduction 483
  The Changing Roles of Data and Database Administrators 484
    Data Administration 484
    Database Administration 485
    Changing Approaches to Data Administration 488
Modeling Enterprise Data 490
Planning for Databases 491
Managing Data Security 492
  Threats to Data Security 492
  Views 494
  Authorization Rules 495
  User-Defined Procedures 497
  Encryption 497
  Authentication Schemes 497
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backing Up Databases</td>
<td>498</td>
</tr>
<tr>
<td>Basic Recovery Facilities</td>
<td>499</td>
</tr>
<tr>
<td>Backup Facilities</td>
<td>499</td>
</tr>
<tr>
<td>Journalizing Facilities</td>
<td>499</td>
</tr>
<tr>
<td>Checkpoint Facility</td>
<td>500</td>
</tr>
<tr>
<td>Recovery Manager</td>
<td>500</td>
</tr>
<tr>
<td>Recovery and Restart Procedures</td>
<td>500</td>
</tr>
<tr>
<td>Switch</td>
<td>501</td>
</tr>
<tr>
<td>Restore/Rerun</td>
<td>501</td>
</tr>
<tr>
<td>Transaction Integrity</td>
<td>501</td>
</tr>
<tr>
<td>Backward Recovery</td>
<td>502</td>
</tr>
<tr>
<td>Forward Recovery</td>
<td>503</td>
</tr>
<tr>
<td>Types of Database Failure</td>
<td>503</td>
</tr>
<tr>
<td>Aborted Transactions</td>
<td>503</td>
</tr>
<tr>
<td>Incorrect Data</td>
<td>504</td>
</tr>
<tr>
<td>System Failure</td>
<td>504</td>
</tr>
<tr>
<td>Database Destruction</td>
<td>505</td>
</tr>
<tr>
<td>Controlling Concurrent Access</td>
<td>505</td>
</tr>
<tr>
<td>The Problem of Lost Updates</td>
<td>505</td>
</tr>
<tr>
<td>Serializability</td>
<td>506</td>
</tr>
<tr>
<td>Locking Mechanisms</td>
<td>507</td>
</tr>
<tr>
<td>Locking Level</td>
<td>507</td>
</tr>
<tr>
<td>Types of Locks</td>
<td>508</td>
</tr>
<tr>
<td>Deadlock</td>
<td>508</td>
</tr>
<tr>
<td>Managing Deadlock</td>
<td>510</td>
</tr>
<tr>
<td>Versioning</td>
<td>510</td>
</tr>
<tr>
<td>Managing Data Quality</td>
<td>511</td>
</tr>
<tr>
<td>Security Policy and Disaster Recovery</td>
<td>512</td>
</tr>
<tr>
<td>Personnel Controls</td>
<td>513</td>
</tr>
<tr>
<td>Physical Access Controls</td>
<td>513</td>
</tr>
<tr>
<td>Maintenance Controls</td>
<td>514</td>
</tr>
<tr>
<td>Data Protection and Privacy</td>
<td>514</td>
</tr>
<tr>
<td>Data Dictionaries and Repositories</td>
<td>514</td>
</tr>
<tr>
<td>Repositories</td>
<td>515</td>
</tr>
<tr>
<td>Overview of Tuning the Database for Performance</td>
<td>516</td>
</tr>
<tr>
<td>Installation of the DBMS</td>
<td>517</td>
</tr>
<tr>
<td>Memory Usage</td>
<td>517</td>
</tr>
<tr>
<td>Input/Output (I/O) Contention</td>
<td>518</td>
</tr>
<tr>
<td>CPU Usage</td>
<td>518</td>
</tr>
<tr>
<td>Application Tuning</td>
<td>518</td>
</tr>
<tr>
<td>Summary</td>
<td>519</td>
</tr>
<tr>
<td>Chapter Review</td>
<td>520</td>
</tr>
<tr>
<td>Key Terms</td>
<td>520</td>
</tr>
<tr>
<td>Review Questions</td>
<td>520</td>
</tr>
<tr>
<td>Problems and Exercises</td>
<td>522</td>
</tr>
<tr>
<td>Field Exercises</td>
<td>525</td>
</tr>
<tr>
<td>References</td>
<td>525</td>
</tr>
<tr>
<td>Project Case: Mountain View Community Hospital</td>
<td>527</td>
</tr>
</tbody>
</table>

Chapter 14: Data Warehouse

Learning Objectives: 529
Introduction 529

Basic Concepts of Data Warehousing 531
  A Brief History 531
  The Need for Data Warehousing 532
    Need for a Company-Wide View 532
    Need to Separate Operational and Information Systems 533

Data Warehouse Architectures 534
  Generic Two-Level Architecture 534
  An Expanded Data Warehouse Architecture 534
  Three-Layer Data Architecture 537
    Role of the Enterprise Data Model 537
    Role of Metadata 537

Some Data Characteristics 538
  Status versus Event Data 538
  Transient versus Periodic Data 539
  An Example of Transient and Periodic Data 540
    Transient Data 541
    Periodic Data 542

The Reconciled Data Layer 542
  Characteristics of Reconciled Data 543
  The Data Reconciliation Process 543
    Capture 544
    Scrub 544
    Load and Index 545

Data Transformation 546
  Data Transformation Functions 547
    Record-Level Functions 547
    Field-Level Functions 548
    More Complex Transformations 550
  Tools to Support Data Reconciliation 550
    Data Quality Tools 550
    Data Conversion Tools 550
    Data-Cleansing Tools 550

The Derived Data Layer 551
  Characteristics of Derived Data 551
  The Star Schema 552
    Fact Tables and Dimension Tables 552
    Example Star Schema 552
    Grain of a Fact Table 554
    Size of the Fact Table 554
  Variations of the Star Schema 556
    Multiple Fact Tables 556
    Snowflake Schema 556
    Proprietary Databases 557

Independent versus Dependent Data Marts 558

The User Interface 559
  Role of Metadata 559
  On-Line Analytical Processing (OLAP) Tools 560
    Slicing a Cube 561
    Drill-Down 561
Data-Mining Tools 562
   Data-Mining Techniques 562
   Data-Mining Applications 562
   Data Visualization 563

Summary 563

Chapter Review 565
   Key Terms 565
   Review Questions 565
   Problems and Exercises 566
   Field Exercises 568
   References 568
   Further Reading 569

Project Case: Mountain View Community Hospital 570

Appendix A 573
   Object-Relational Databases
   Basic Concepts and Definitions 573
   Features of an ORDBMS 574
   Complex Data Types 574
   Enhanced SQL 575
   A Simple Example 575
   Content Addressing 576
   Advantages of the Object-Relational Approach 576
   ORDBMS Vendors and Products 577
   References 577

Appendix B 579
   Advanced Normal Forms
   Boyce-Codd Normal Form 579
   Anomalies in STUDENT_ADVISOR 579
   Definition of Boyce-Codd Normal Form (BCNF) 580
   Converting a Relation to BCNF 580
   Fourth Normal Form 582
   Multivalued Dependencies 583
   Higher Normal Forms 584
   References 584

Appendix C 585
   Data Structures
   Pointers 585
   Data Structure Building Blocks 587
   Linear Data Structures 589
   Stacks 590
   Queues 590
   Sorted Lists 591
   Multilists 593
   Hazards of Chain Structures 594
   Trees 594
   Balanced Trees 595
   References 598
Glossary of Terms  599
Glossary of Acronyms  611
Index  613