Modern Database Management

NINTH EDITION

Jeffrey A. Hoffer
University of Dayton

Mary B. Prescott
University of Tampa

Heikki Topi
Bentley College

Pearson Education International

CONTENTS

Preface 29

PART I THE CONTEXT OF DATABASE MANAGEMENT 41

An Overview of Part One 41

CHAPTER 1 The Database Environment 43

Learning Objectives 43

Data Matters! 43

Introduction 45

Basic Concepts and Definitions 46

Data 46

Data Versus Information 47

Metadata 48

Database Management Systems 49

Data Models 49

Entities 49

Relationships 49

Relational Databases 51

Traditional File Processing Systems 51

File Processing Systems at Pine Valley Furniture Company 51

Disadvantages of File Processing Systems 52

Program-Data Dependence 52

Duplication of Data 52

Limited Data Sharing 53

Lengthy Development Times 53

Excessive Program Maintenance 53

The Database Approach 53

Advantages of The Database Approach 54

Program-Data Independence 54

Planned Data Redundancy 54

Improved Data Consistency 55

Improved Data Sharing 55

Increased Productivity of Application Development 55

Enforcement of Standards 55

Improved Data Quality 56

Improved Data Accessibility and Responsiveness 56

Reduced Program Maintenance 56

Improved Decision Support 57

Cautions About Database Benefits 57

Costs and Risks of The Database Approach 57

New, Specialized Personnel 57

Installation and Management Cost and Complexity 57

Conversion Costs 58

Need for Explicit Backup and Recovery 58

Organizational Conflict 58

Components of the Database Environment 58

The Range of Database Applications 60

Personal Databases 60

Workgroup Databases 61

Departmental/Divisional Databases 63
Enterprise Databases 63
Web-Enabled Databases 65
Summary of Database Applications 66
Evolution of Database Systems 67
1960s 67
1970s 68
1980s 68
1990s 68
2000 and Beyond 68
Summary 69
Chapter Review 70
Key Terms 70
Review Questions 70
Problems and Exercises 71
Field Exercises 73
References 74
Further Reading 74
Web Resources 74
Case: Mountain View Community Hospital 75
CHAPTER 2 The Database Development Process 78
Learning Objectives 78
Introduction 79
Database Development Within Information Systems Development 80
Information Systems Architecture 80
Information Engineering 81
Information Systems Planning 81
Identifying Strategic Planning Factors 81
Identifying Corporate Planning Objects 82
Developing an Enterprise Model 82
Database Development Process 85
Systems Development Life Cycle 85
Planning–Enterprise Modeling 86
Planning–Conceptual Data Modeling 87
Analysis–Conceptual Data Modeling 88
Design–Logical Database Design 88
Design–Physical Database Design and Definition 89
Implementation–Database Implementation 89
Maintenance–Database Maintenance 89
Alternative IS Development Approaches 89
Role of Packaged Data Models 91
Universal Data Models 92
Industry-Specific Data Models 92
Summary 92
The Role of CASE and a Repository 93
Managing the People Involved in Database Development 94
Three-Schema Architecture for Database Development 95
Three-Schema Components 95
Summary of Schemas 96
Enterprise Data Model 97
User Views 97

Conceptual Schema 97
Logical Schema 97
Physical Schema 98
Strategies for Development 98
Three-Tiered Database Location Architecture 99
Developing a Database Application for Pine Valley Furniture Company 101
Simplified Project Data Model Example 102
A Current Pine Valley Furniture Company Project Request 105
Matching User Needs to the Information Systems Architecture 106
Analyzing Database Requirements 108
Designing the Database 109
Using the Database 111
Administering the Database 114
Summary 114
Chapter Review 115
Key Terms 115
Review Questions 115
Problems and Exercises 116
Field Exercises 118
References 119
Further Reading 119
Web Resources 119
Case: Mountain View Community Hospital 120

PART II DATABASE ANALYSIS 127

An Overview of Part Two **127**

CHAPTER 3 Modeling Data in the Organization 129

Learning Objectives 129
Introduction 130
Modeling the Rules of the Organization 131
Overview of Business Rules 132
The Business Rules Paradigm 132
Scope of Business Rules 133
Good Business Rules 133
Gathering Business Rules 133
Data Names and Definitions 134
Data Names 134
Data Definitions 135
Good Data Definitions 136
The E-R Model: An Overview 137
Sample E-R Diagram 137
E-R Model Notation 139
Modeling Entities and Attributes 140
Entities 140
Entity Type Versus Entity Instance 141
Entity Type Versus System Input, Output, or User 141
Strong Versus Weak Entity Types 142
Naming and Defining Entity Types 143
Attributes 145
Required Versus Optional Attributes 145
Simple Versus Composite Attributes 146

Single-Valued Versus Multivalued Attributes 146
Stored Versus Derived Attributes 147
Identifier Attribute 147
Naming and Defining Attributes 149
Modeling Relationships 150
Basic Concepts and Definitions in Relationships 151
Attributes on Relationships 152
Associative Entities 152
Degree of a Relationship 154
Unary Relationship 154
Binary Relationship 157
Ternary Relationship 157
Attributes or Entity? 158
Cardinality Constraints 160
Minimum Cardinality 160
Maximum Cardinality 161
Some Examples 161
A Ternary Relationship 162
Modeling Time-Dependent Data 163
Multiple Relationships 165
Naming and Defining Relationships 167
E-R Modeling Example: Pine Valley Furniture Company 168
Database Processing at Pine Valley Furniture 171
Showing Product Information 171
Showing Customer Information 172
Showing Customer Order Status 172
Showing Product Sales 173
Summary 174
Chapter Review 175
Key Terms 175
Review Questions 176
Problems and Exercises 177
Field Exercises 186
References 186
Further Reading 187
Web Resources 187
Case: Mountain View Community Hospital 188
CHAPTER 4 The Enhanced E-R Model and Business Rules 190
Learning Objectives 190
Introduction 191
Representing Supertypes and Subtypes 192
Basic Concepts and Notation 192
An Example 194
Attribute Inheritance 194
When to Use Supertype/Subtype Relationships 195
Representing Specialization and Generalization 196
Generalization 196
Specialization 197
Combining Specialization and Generalization 198
Specifying Constraints in Supertype/Subtype Relationships 199

Specifying Completeness Constraints 199
Total Specialization Rule 199
Partial Specialization Rule 199
Specifying Disjointness Constraints 199
Disjoint Rule 200
Overlap Rule 201
Defining Subtype Discriminators 202
Disjoint Subtypes 202
Overlapping Subtypes 203
Defining Supertype/Subtype Hierarchies 203
An Example 204
Summary of Supertype/Subtype Hierarchies 205
EER Modeling Example: Pine Valley Furniture 205
Entity Clustering 209
Packaged Data Models 212
Business Rules Revisited 218
Classification of Business Rules 219
Stating a Structural Assertion 220
Derived Facts 221
Stating an Action Assertion 221
Types of Action Assertions 222
Representing and Enforcing Business Rules 222
Sample Business Rules 223
Identifying and Testing Business Rules 225
Summary 226
Chapter Review 227
Key Terms 227
Review Questions 227
Problems and Exercises 228
Field Exercises 234
References 234
Further Reading 235
Web Resources 235
Case: Mountain View Community Hospital 236

PART III DATABASE DESIGN 239
An Overview of Part Three 239

CHAPTER 5 Logical Database Design and the Relational Model 241
Learning Objectives 241
Introduction 241
The Relational Data Model 242
Basic Definitions 243
Relational Data Structure 243
Relational Keys 243
Properties of Relations 244
Removing Multivalued Attributes from Tables 244
Example Database 245
Integrity Constraints 247
Domain Constraints 247
Entity Integrity 248

Referential Integrity 248
Action Assertions 249
Creating Relational Tables 250
Well-Structured Relations 251
Transforming EER Diagrams into Relations 252
Step 1: Map Regular Entities 252
Composite Attributes 253
Multivalued Attributes 253
Step 2: Map Weak Entities 254
When to Create a Surrogate Key 255
Step 3: Map Binary Relationships 256
Map Binary One-to-Many Relationships 256
Map Binary Many-to-Many Relationships 257
Map Binary One-to-One Relationships 257
Step 4: Map Associative Entities 258
Identifier Not Assigned 259
Identifier Assigned 259
Step 5: Map Unary Relationships 261
Unary One-to-Many Relationships 261
Unary Many-to-Many Relationships 261
Step 6: Map Ternary (and n-ary) Relationships 262
Step 7: Map Supertype/Subtype Relationships 264
Summary of EER to Relational Transformations 265
Introduction to Normalization 266
Steps in Normalization 267
Functional Dependencies and Keys 268
Determinants 268
Candidate Keys 268
Normalization Example: Pine Valley Furniture Company 270
Step 0: Represent the View in Tabular Form 270
Step 1: Convert to First Normal Form 271
Remove Repeating Groups 271
Select the Primary Key 272
Anomalies in 1NF 272
Step 2: Convert to Second Normal Form 273
Step 3: Convert to Third Normal Form 274
Removing Transitive Dependencies 274
Determinants and Normalization 275
Step 4: Further Normalization 276
Merging Relations 276
An Example 277
View Integration Problems 277
Synonyms 277
Homonyms 277
Transitive Dependencies 278
Supertype/Subtype Relationships 278
A Final Step for Defining Relational Keys 279
Summary 280
Chapter Review 282
Key Terms 282
Review Questions 282
Problems and Exercises 283
Field Exercises 290

References 291
Further Reading 291
Web Resources 291
Case: Mountain View Community Hospital 292

CHAPTER 6 Physical Database Design and Performance 297
Learning Objectives 297
Introduction 297
Physical Database Design Process 298
Data Volume and Usage Analysis 299
Designing Fields 301
Choosing Data Types 301
Coding and Compression Techniques 302
Controlling Data Integrity–A Foundation for Sarbanes-Oxley Compliance 303
Handling Missing Data 305
Designing Physical Records and Denormalization 305
Denormalization 306
Denormalize with Caution 306
Opportunities and Types of Denormalization 307
Designing Physical Files 312
Pointer 313
File Organizations 313
Sequential File Organizations 314
Indexed File Organizations 314
Hashed File Organizations 319
Summary of File Organizations 321
Clustering Files 321
Designing Controls for Files 323
Using and Selecting Indexes 323
Creating a Unique Key Index 323
Creating a Secondary (Nonunique) Key Index 324
When to Use Indexes 324
RAID: Improving File Access Performance by Parallel Processing 325
Designing Databases 328
Optimizing for Query Performance 331
Parallel Query Processing 331
Overriding Automatic Query Optimization 332
Picking Data Block Size 333
Balancing I/O Across Disk Controllers 334
Guidelines for Better Query Design 334
Summary 336
Chapter Review 337
Key Terms 337
Review Questions 338
Problems and Exercises 338
Field Exercises 342
References 342
Further Reading 343
Web Resources 343
Case: Mountain View Community Hospital 344

PART IV IMPLEMENTATION 347

An Overview of Part Four 347

CHAPTER 7 Introduction to SQL 349

Learning Objectives 349

Introduction 349

History of the SQL Standard 351

The Role of SQL in a Database Architecture 352

The SQL Environment 353

Defining a Database in SQL 358

Generating SQL Database Definitions 358

Creating Tables 359

Creating Data Integrity Controls 362

Changing Table Definitions 363

Removing Tables 364

Inserting, Updating, and Deleting Data 364

Batch Input 366

Deleting Database Contents 366

Updating Database Contents 366

Internal Schema Definition in RDBMSs 367

Creating Indexes 368

Processing Single Tables 368

Clauses of the SELECT Statement 369

Using Expressions 371

Using Functions 371

Using Wildcards 372

Using Comparison Operators 373

Using Boolean Operators 373

Using Ranges for Qualification 376

Using Distinct Values 377

Using IN and NOT IN with Lists 379

Sorting Results: The ORDER BY Clause 379

Categorizing Results: The GROUP BY Clause 380

Qualifying Results by Categories: The HAVING Clause 381

Using and Defining Views 382

Materialized Views 386

Summary 387

Chapter Review 387

Key Terms 387

Review Questions 387

Problems and Exercises 389

Field Exercises 393

References 394

Further Reading 394

Web Resources 394

Case: Mountain View Community Hospital 395

CHAPTER 8 Advanced SQL 396

Learning Objectives 396

Introduction 396

Processing Multiple Tables 397
Equi-join 398
Natural Join 400
Outer Join 401
Union Join 402
Sample Multiple Join Involving Four Tables 403
Subqueries 404
Correlated Subqueries 408
Using Derived Tables 409
Combining Queries 409
Conditional Expressions 412
More Complicated SQL Queries 413
Tips for Developing Queries 415
Ensuring Transaction Integrity 416
Data Dictionary Facilities 418
SQL:200N Enhancements and Extensions to SQL 420
Analytical Functions 420
New Data Types 421
Other Enhancements 422
Programming Extensions 422
Triggers and Routines 424
Triggers 425
Routines 426
Embedded SQL and Dynamic SQL 428
OLAP SQL 431
Summary 433
Chapter Review 434
Key Terms 434
Review Questions 434
Problems and Exercises 435
Field Exercises 439
References 439
Further Reading 439
Web Resources 439
Case: Mountain View Community Hospital 440

CHAPTER 9 The Client/Server Database Environment 441
Learning Objectives 441
Location, Location, Location! 441
Introduction 443
Client/Server Architectures 444
File Server Architectures 444
Limitations of File Servers 445
Database Server Architectures 446
Three-Tier Architectures 447
Partitioning an Application 449
Role of the Mainframe 452
Using Middleware 453
Client/Server Issues 455

Using ODBC to Link External Tables Stored on a Database Server 457
Using JDBC to Link External Tables Stored on a Database Server 459
Looking Forward with Client/Server in Mind 459
Summary 459
Chapter Review 460
Key Terms 460
Review Questions 461
Problems and Exercises 461
Field Exercises 462
References 462
Further Reading 462
Web Resources 463
Case: Mountain View Community Hospital 464

CHAPTER 10 The Internet Database Environment 466
Learning Objectives 466
Introduction 467
The Internet and Database Connection 467
The Internet Environment 468
Common Internet Architecture Components 470
Internet-Related Languages 470
XML and XQuery Overview 473
Server-Side Extensions 477
Web Server Interfaces 478
Web Servers 479
Client-Side Extensions 480
Web-to-Database Tools 481
Web Services 483
Lack of Mature Standards 490
Lack of Security 490
Service-Oriented Architecture (SOA) 491
Semantic Web 491
Internet Technology Rate-of-Change Issues 491
Summary 492
Chapter Review 493
Key Terms 493
Review Questions 493
Problems and Exercises 494
Field Exercises 495
References 495
Further Reading 495
Web Resources 496
Case: Mountain View Community Hospital 497

CHAPTER 11 Data Warehousing 499
Learning Objectives 499
Introduction 500
Basic Concepts of Data Warehousing 502

A Brief History 503
The Need for Data Warehousing 503
Need for a Company-wide View 503
Need to Separate Operational and Informational Systems 505
Data Warehousing Success 506
Data Warehouse Architectures 507
Independent Data Mart Data Warehousing Environment 508
Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach 510
Logical Data Mart and Real-Time Data Warehouse Architecture 512
Three-Layer Data Architecture 515
Role of the Enterprise Data Model 515
Role of Metadata 516
Some Characteristics of Data Warehouse Data 516
Status Versus Event Data 516
Transient Versus Periodic Data 517
An Example of Transient and Periodic Data 517
Transient Data 519
Periodic Data 519
Other Data Warehouse Changes 519
The Derived Data Layer 520
Characteristics of Derived Data 520
The Star Schema 521
Fact Tables and Dimension Tables 521
Example Star Schema 522
Surrogate Key 524
Grain of Fact Table 525
Duration of the Database 526
Size of the Fact Table 526
Modeling Date and Time 527
Variations of the Star Schema 528
Multiple Fact Tables 528
Factless Fact Tables 529
Normalizing Dimension Tables 530
Multivalued Dimensions 530
Hierarchies 531
Slowly Changing Dimensions 533
Determining Dimensions and Facts 536
The User Interface 538
Role of Metadata 538
SQL OLAP Querying 538
Online Analytical Processing (OLAP) Tools 540
Slicing a Cube 541
Drill-Down 542
Summarizing More Than Three Dimensions 543
Data Visualization 543
Business Performance Management and Dashboards 544
Data-Mining Tools 545
Data-Mining Techniques 545
Data-Mining Applications 546
Summary 547
Chapter Review 548
Key Terms 548
Review Questions 548
Problems and Exercises 549

Field Exercises 554
References 554
Further Reading 555
Web Resources 555
Case: Mountain View Community Hospital 556

PART V ADVANCED DATABASE TOPICS 559
An Overview of Part Five 559

CHAPTER 12 Data Quality and Integration 562
Learning Objectives 562
Introduction 562
Managing Data Quality 564
The State of Data Quality 566
External Data Sources 567
Redundant Data Storage and Inconsistent Metadata 567
Data Entry 567
Lack of Organizational Commitment 567
Data Quality Improvement 567
Conduct a Data Quality Audit 567
Improve Data Capture Processes 569
Establish a Data Stewardship Program 569
Apply TQM Principles and Practices 571
Apply Modern Data Management Technology 571
Estimate Return on Investment 571
Start with a High-Quality Data Model 571
Summary of Data Quality 573
Data Integration: An Overview 573
General Approaches to Data Integration 573
Data Federation 574
Data Propagation 574
Master Data Management 575
Data Integration for Data Warehousing: The Reconciled Data Layer 577
Characteristics of Data After ETL 577
The ETL Process 578
Mapping and Metadata Management 579
Extract 579
Cleanse 580
Load and Index 583
Data Transformation 584
Data Transformation Functions 585
Record-Level Functions 585
Field-Level Functions 586
More Complex Transformations 586
Tools to Support Data Reconciliation 586
Data Quality Tools 587
Data Conversion Tools 589
Data Cleansing Tools 589
Selecting Tools 589
Summary 589
Chapter Review 590
Key Terms 590

Review Questions 590
Problems and Exercises 591
Field Exercises 592
References 592
Further Reading 593
Web Resources 593
Case: Mountain View Community Hospital 594

CHAPTER 13 Data and Database Administration 596
Learning Objectives 596
Introduction 597
The Roles of Data and Database Administrators 598
Traditional Data Administration 599
Traditional Database Administration 600
Evolving Approaches to Data Administration 603
Blending Data and Database Administration 603
Fast-Track Development 603
New DBA Roles 603
Summary of Evolving Data Administration Roles 605
The Open-Source Movement 605
Modeling Enterprise Data 607
Organizational Roles 607
Role of an Information Systems Architecture 608
Managing Data Security 608
Threats to Data Security 609
Establishing Client/Server Security 610
Server Security 611
Network Security 611
Client/Server Security Issues for Web-Enabled Databases 611
Web Security 612
Web Privacy 613
Database Software Data Security Features 614
Views 614
Integrity Controls 615
Authorization Rules 617
User-Defined Procedures 618
Encryption 619
Authentication Schemes 620
Passwords 620
Strong Authentication 621
Mediated Authentication 622
Security Policies and Procedures 622
Personnel Controls 622
Physical Access Controls 623
Maintenance Controls 623
Data Privacy Controls 623
Database Backup and Recovery 624
Basic Recovery Facilities 624
Backup Facilities 624
Journalizing Facilities 625
Checkpoint Facility 626
Recovery Manager 626
Recovery and Restart Procedures 626

Disk Mirroring 626
Restore/Rerun 626
Maintaining Transaction Integrity 627
Backward Recovery 628
Forward Recovery 629
Types of Database Failure 630
Aborted Transactions 630
Incorrect Data 630
System Failure 631
Database Destruction 631
Disaster Recovery 631
Controlling Concurrent Access 632
The Problem of Lost Updates 632
Serializability 633
Locking Mechanisms 633
Locking Level 634
Types of Locks 635
Deadlock 636
Managing Deadlock 636
Versioning 637
Data Dictionaries and Repositories 638
Data Dictionary 638
Repositories 639
Overview of Tuning the Database for Performance 641
Installation of the DBMS 641
Memory and Storage Space Usage 642
Input/Output (I/O) Contention 642
CPU Usage 643
Application Tuning 643
Data Availability 644
Costs of Downtime 644
Measures to Ensure Availability 645
Hardware Failures 645
Loss or Corruption of Data 645
Human Error 645
Maintenance Downtime 646
Network-Related Problems 646
Summary 646
Chapter Review 647
Key Terms 647
Review Questions 647
Problems and Exercises 649
Field Exercises 652
References 653
Further Reading 653
Web Resources 653
Case: Mountain View Community Hospital 654
CHAPTER 14 Overview: Distributed Databases 655
Learning Objectives 655
Overview 656
Objectives and Trade-Offs 656
Options for Distributing a Database 656

Distributed DBMS 657
Query Optimization 658
Chapter Review 658
References 658
Further Reading 659
Web Resources 659

CHAPTER 15 Overview: Object-Oriented Data Modeling 660
Learning Objectives 660
Overview 661
The Unified Modeling Language 661
Object-Oriented Data Modeling 662
Representing Aggregation 667
Chapter Review 668
References 668
Further Reading 668
Web Resources 668

CHAPTER 16 Overview: Using Relational Databases to Provide Object Persistence 669
Learning Objectives 669
Overview 670
Providing Persistence for Objects Using Relational Databases 671
Call-Level Application Program Interfaces 671
SQL Query Mapping Frameworks 671
Object-Relational Mapping Frameworks 671
Proprietary Approaches 672
Object-Relational Mapping Example 673
Mapping Files 674
Responsibilities of Object-Relational Mapping Frameworks 676
Summary 677
Chapter Review 678
References 678
Further Reading 679
Web Resources 679

APPENDIX A Data Modeling Tools and Notation 680
Comparing E-R Modeling Conventions 680
Visio Professional 2003 Notation 683
Entities 683
Relationships 683
AllFusion ERwin Data Modeler 4.1 SP1 Notation 683
Entities 683
Relationships 685
Sybase Power Designer 11.1 Notation 685
Entities 685
Relationships 687
Oracle Designer Notation 687
Entities 687
Relationships 688
Comparison of Tool Interfaces and E-R Diagrams 688

APPENDIX B Advanced Normal Forms 691
Boyce-Codd Normal Form 691

Anomalies in STUDENT_ADVISOR 691
Definition of Boyce-Codd Normal Form (BCNF) 692
Converting a Relation to BCNF 692
Fourth Normal Form 694
Multivalued Dependencies 695
Higher Normal Forms 696
Appendix Review 696
Key Terms 696
References 696
Web Resources 696

APPENDIX C Data Structures 697
Pointers 697
Data Structure Building Blocks 699
Linear Data Structures 699
Stacks 702
Queues 702
Sorted Lists 702
Multilists 705
Hazards of Chain Structures 706
Trees 706
Balanced Trees 707
References 709
Glossary of Acronyms 710
Glossary of Terms 712
Index 721