### CONTENTS

#### Introduction

- Level of Understanding  
- How to Use the Book Actively  
- The Purpose of Each Chapter
  - *Section 1: Project Management and Requirements*  
  - *Section 2: Data Design*  
  - *Section 3: Architecture*  
  - *Section 4: Implementation*  
  - *Section 5: Deployment and Growth*  
  - *Supporting Tools*  
- The Goals of a Data Warehouse  
- The Goals of this Book  
- Visit the Companion Web Site  

#### Chapter 1  The Chess Pieces

- Basic Elements of the Data Warehouse
  - *Source System*  
  - *Data Staging Area*  
  - *Presentation Server*  
  - *Dimensional Model*  
  - *Business Process*  
  - *Data Mart*  
  - *Data Warehouse*  
  - *Operational Data Store (ODS)*  
  - *OLAP (On-Line Analytic Processing)*  
  - *ROLAP (Relational OLAP)*  
  - *MOLAP (Multidimensional OLAP)*
Chapter 4 Collecting the Requirements 95

Overall Approach to Requirements Definition 96
  Interviews versus Facilitated Sessions 97
Prepare for the Interview 98
  Identify the Interview Team 98
  Conduct the Pre-interview Research 101
  Select the Interviewees 102
  Develop the Interview Questionnaires 104
  Schedule the Interviews 104
  Prepare the Interviewees 108
Conduct the Interview 111
  Remember your Interview Role 111
  Verify Communications 112
  Define Terminology 112
  Establish Peer Basis 113
  Maintain Interview Schedule Flexibility 113
  Avoid Interview Burnout 113
  Manage Expectations Continuously 114
Start the Interview 114
  Business Executive Interview Content 116
  Business Manager or Analyst Interview Content 117
  IS Data Audit Interview Content 120
Wrap Up the Interview 122
  Determine the Success Criteria 122
Review the Interview Results 126
Prepare and Publish the Requirements Deliverables 128
Agree on Next Steps After Collecting Requirements 131
  Facilitation for Confirmation, Prioritization, and Consensus 132
  Facilitation as an Alternative to Interviewing 132
Summary 133

SECTION 2 DATA DESIGN 137

Chapter 5 A First Course on Dimensional Modeling 139

The Case for Dimensional Modeling 140
  What Is Entity-Relationship Modeling? 140
  What Is Dimensional Modeling? 144
The Relationship between Dimensional Modeling and Entity-Relationship Modeling 146
The Strengths of Dimensional Modeling 147
Putting Dimensional Models Together: The Data Warehouse Bus Architecture 153
The Planning Crisis 153
Data Marts with a Bus Architecture 155
Conformed Dimensions and Standard Fact Definitions 156
Designing the Conformed Dimensions 158
Taking the Pledge 158
Establishing the Conformed Fact Definitions 159
The Importance of Data Mart Granularity 160
Multiple-Source Data Marts 161
Rescuing Stovepipes 162
When You Don't Need Conformed Dimensions 163
The Data Warehouse Bus 164
Basic Dimensional Modeling Techniques 164
Fact Tables and Dimension Tables 165
Inside Dimension Tables, Drilling Up and Down 167
Foreign Keys, Primary Keys, and Surrogate Keys 191
Additive, Semiadditive, and Nonadditive Facts 193
The Four-Step Design Method for Designing an Individual Fact Table 194
Families of Fact Tables 200
Factless Fact Tables 212
Summary 216

Chapter 6 A Graduate Course on Dimensional Modeling 217

Extended Dimension Table Designs 218
Many-to-Many Dimensions 218
Many-to-One-to-Many Traps 222
Role-Playing Dimensions 223
Organization and Parts Hierarchies 226
Unpredictably Deep Hierarchies 231
Time Stamping the Changes in a Large Dimension 233
Building an Audit Dimension 237
Too Few Dimensions and Too Many Dimensions 238
Extended Fact Table Designs 240
Facts of Differing Granularity and Allocating 240
Time of Day 245
Multiple Units of Measure 247
Multinational Currency Tracking 248
Value Band Reporting 251
Advanced ROLAP Querying and Reporting 252
Drill-Across Queries with Multiple Technologies 252
Self Referencing Queries, Behavior Tracking, and Sequential Subsetting 254
Market Basket Analysis 260
Summary 263

Chapter 7 Building Dimensional Models 265
Matrix Method for Getting Started 266
Build the Matrix 266
Use the Four-Step Method to Design Each Fact Table 272
Managing the Dimensional Modeling Project 276
Data Warehouse Bus Architecture Matrix 277
Fact Table Diagram 277
Fact Table Detail 280
Dimension Table Detail 281
Steps for the Dimensional Modeling Team 285
Managing Issues in the Modeling Process 294
Identifying the Sources for Each Fact Table and Dimension Table 296
When Are You Done? 307
Using a Data Modeling Tool 309
Summary 312

SECTION 3 ARCHITECTURE 315

Chapter 8 Introducing Data Warehouse Architecture 317
The Value of Architecture 318
An Architectural Framework and Approach 320
Defining the Columns 322
Defining the Levels of Detail (the Rows) 324
Logical Models and Physical Models 326
Framework Summary 327
Technical Architecture Overview 327
Services and Data Stores 330
Flow from Source System to User Desktop 330
Key Technical Architecture Features 331
Evolution of Your Data Warehouse Architecture 333
Summary 334

Chapter 9 Back Room Technical Architecture 335
Back Room Data Stores 336
Source Systems 336
Data Staging Area 344
The Presentation Servers 346
<table>
<thead>
<tr>
<th>Chapter 10</th>
<th>Architecture for the Front Room</th>
<th>373</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front Room Data Stores</td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>Access Tool Data Stores</td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>Standard Reporting Data Stores</td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>Downstream Systems</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td>Front Room Services for Data Access</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td>Warehouse Browsing</td>
<td>379</td>
<td></td>
</tr>
<tr>
<td>Access and Security Services</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Activity Monitoring Services</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Query Management Services</td>
<td>381</td>
<td></td>
</tr>
<tr>
<td>Query Service Locations</td>
<td>385</td>
<td></td>
</tr>
<tr>
<td>Standard Reporting Services</td>
<td>386</td>
<td></td>
</tr>
<tr>
<td>Future Access Services</td>
<td>388</td>
<td></td>
</tr>
<tr>
<td>Desktop Services</td>
<td>389</td>
<td></td>
</tr>
<tr>
<td>Modeling Applications and Data Mining</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>Web Implications for Data Access</td>
<td>404</td>
<td></td>
</tr>
<tr>
<td>Desktop Tool Architecture Approaches</td>
<td>406</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>409</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 11</th>
<th>Infrastructure and Metadata</th>
<th>411</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>412</td>
<td></td>
</tr>
<tr>
<td>Drivers of Infrastructure</td>
<td>412</td>
<td></td>
</tr>
<tr>
<td>The Evolution of Infrastructure</td>
<td>413</td>
<td></td>
</tr>
<tr>
<td>Back Room Infrastructure Factors</td>
<td>413</td>
<td></td>
</tr>
<tr>
<td>Front Room Infrastructure Factors</td>
<td>429</td>
<td></td>
</tr>
<tr>
<td>Connectivity and Networking Factors</td>
<td>432</td>
<td></td>
</tr>
<tr>
<td>Infrastructure Summary</td>
<td>434</td>
<td></td>
</tr>
<tr>
<td>Metadata and the Metadata Catalog</td>
<td>435</td>
<td></td>
</tr>
<tr>
<td>Metadata: What Is It?</td>
<td>435</td>
<td></td>
</tr>
<tr>
<td>Source System Metadata</td>
<td>436</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 12  A Graduate Course on the Internet and Security 449

Component Software Architecture 452
  The Impact of the Internet on Architecture 454
Security: Vulnerabilities 457
  Physical Assets 459
  Information Assets: Data, Financial Assets, and Reputation 459
  Software Assets 467
  The Ability to Conduct Your Business 467
  Network Threats 467
Security: Solutions 472
  Routers and Firewalls 474
  The Directory Server 479
  Encryption 481
Managing Security in a Data Warehouse Environment 495
What to do Now About Security 497
  Immediate Tactical Measures 497
  Strategic Measures 498
Summary 499

Chapter 13  Creating the Architecture Plan and Selecting Products 501

Creating the Architecture 502
  Architecture Development Process 503
Selecting the Products 514
  Keeping a Business Focus 514
  Major Warehouse Evaluation Areas 515
  Evaluation Process 516
  Evaluations for the Back Room 530
  Evaluations for the Front Room 532
Installation 535
Summary 536

SECTION 4  IMPLEMENTATION 541

Chapter 14  A Graduate Course on Aggregates 543

Aggregation Goals and Risks 544
Deciding What to Aggregate 545
Develop the Aggregate Table Plan 547
Processing Aggregates 549
Administering the Aggregates 552
Design Goals for an Aggregate Navigation System 555
  Design Goal 1 556
  Design Goal 2 557
  Design Goal 3 559
  Design Goal 4 560
 Aggregate Navigation Algorithm 560
  Intersecting Planning Data 564
  Handling MIN, MAX, COUNT, and AVG 565
  Aggregates for Everyone 566
Summary 567

Chapter 15 Completing the Physical Design 569
Develop Standards 571
  Database Object Naming Standards 571
  Use Synonyms for All Tables That Users Access 574
  Physical File Locations 574
Develop the Physical Data Model 575
  The Beverage Store Sales Case Study 577
  Use a Data Modeling Tool 578
  Design the Physical Data Structure 578
  Develop Initial Sizing Estimates 583
Develop the Initial Index Plan 585
  Overview of Indexes and Query Strategies 587
  Indexing Fact Tables 590
  Indexing Dimension Tables 591
  Indexing for Loads 592
  Analyze Tables and Indexes after the Load 592
  Case Study Indexes 592
Design and Build the Database Instance 592
  Memory 595
  Blocksize 595
  Disk Access Example 595
  Save the Database Build Scripts and Parameter Files 597
Develop the Physical Storage Structure 598
  Compute Table and Index Sizes 598
  Develop the Partitioning Plan 598
  Set up RAID 599
Implement Usage Monitoring 601
  Performance 602
  User Support 602
  Marketing 603
  Planning 603
  Tools 604
Summary 605
Chapter 16  Data Staging  609

Data Staging Overview  610
Do the Preliminary Work  611
   Importance of Good System Development Practices  611
Plan Effectively  611
   Step 1. High-Level Plan  612
   Step 2. Data Staging Tools  612
   Step 3. Detailed Plan  615
Dimension Table Staging  617
   Step 4. Populate a Simple Dimension Table  617
   Step 5. Implement Dimension Change Logic  625
   Step 6. Populate Remaining Dimensions  629
Fact Table Loads and Warehouse Operations  630
   Step 7. Historical Load of Atomic-Level Facts  631
   Step 8. Incremental Fact Table Staging  641
   Step 9. Aggregate Table and MOLAP Loads  647
   Step 10. Warehouse Operation and Automation  650
Data Quality and Cleansing  653
   Data Improvement  654
   Data Quality Assurance  658
Miscellaneous Issues  659
   Archiving in the Data Staging Area  659
   Source System Rollback Segments  659
   Disk Space Management  660
Summary  661

Chapter 17  Building End User Applications  665

Role of the End User Application  666
   What Is an End User Application Template?  667
       Lifecycle Timing  669
Application Specification  670
   Determine the Initial Template Set  670
   Design the Template Navigation Approach  674
   Determine Template Standards  675
   Detailed Template Specification  677
   User Review  677
End User Application Development  678
   Select an Implementation Approach  680
   Application Development  681
   Testing and Data Verification  683
   Document and Roll Out  684
   Think Like a Software Development Manager  684
   Application Template Maintenance  685
Summary  686
CONTENTS

SECTION 5  DEPLOYMENT AND GROWTH 689

Chapter 18  Planning the Deployment 691

Determine Desktop Installation Readiness 692
Develop the End User Education Strategy 693
   Integrate and Tailor Education Content 693
   Develop the End User Education 696
   Deliver the End User Education 697
Develop an End User Support Strategy 699
   Determine the Support Organization Structure 700
   Anticipate Data Reconciliation Support 701
   Anticipate End User Application Support 702
   Establish Support Communication and Feedback 702
   Provide Support Documentation 703
Develop the Deployment Release Framework 705
   Alpha Release 706
   Beta Release 706
   Production Release (a.k.a. Deployment) 708
Document Your Deployment Strategy 710
Summary 711

Chapter 19  Maintaining and Growing the Data Warehouse 715

Manage the Existing Data Warehouse Environment 716
   Focus on Business Users 718
   Manage Data Warehouse Operations 720
   Measure and Market Your Data Warehouse Success 723
   Communicate Constantly 726
Prepare for Data Warehouse Growth and Evolution 727
   Establish a Data Warehouse Steering Committee 728
   Prioritize Growth and Evolution Opportunities 729
   Manage Iterative Growth and Evolution Using the Lifecycle 731
Summary 733

Appendix A  Lifecycle Project Plan 737

Appendix B  About the CD-ROM 747

Project Plan 747
Data Warehouse Bus Architecture Example 748
Templates, Checklists, Figures, and Aids 749
   Chapter 2: Business Dimensional Lifecycle 750
   Chapter 3: Project Planning and Management 750
   Chapter 4: Collecting the Requirements 750
   Chapter 7: Building Dimensional Models 751
   Chapter 8: Introducing Data Warehouse Architecture 751
   Chapter 9: Back Room Technical Architecture 751
Chapter 10: Architecture for the Front Room 751
Chapter 11: Infrastructure and Metadata 752
Chapter 12: A Graduate Course on the Internet and Security 752
Chapter 13: Creating the Architecture Plan and Selecting Products 752
Chapter 15: Completing the Physical Design 752
Chapter 16: Data Staging 753
Chapter 17: Building End User Applications 753
Chapter 18: Planning the Deployment 753
Using the CD-ROM 753
CD-ROM Browser Instructions 754
Software Requirements 754

Index 755