



*The Data Warehouse*

# Lifecycle Toolkit

Second Edition

Practical Techniques for  
Building Data Warehouse  
and Business Intelligence  
Systems

**Ralph Kimball**

Margy Ross

Warren Thornthwaite

Joy Mundy

Bob Becker





# **The Data Warehouse Lifecycle Toolkit**

**Second Edition**





# **The Data Warehouse Lifecycle Toolkit**

---

## **Second Edition**

Ralph Kimball  
Margy Ross  
Warren Thornthwaite  
Joy Mundy  
Bob Becker



Wiley Publishing, Inc.

## **The Data Warehouse Lifecycle Toolkit, Second Edition**

Published by  
**Wiley Publishing, Inc.**  
10475 Crosspoint Boulevard  
Indianapolis, IN 46256  
[www.wiley.com](http://www.wiley.com)

Copyright © 2008 by Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy and Bob Becker

Published by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-0-470-14977-5

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, or online at <http://www.wiley.com/go/permissions>.

**Limit of Liability/Disclaimer of Warranty:** The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services or to obtain technical support, please contact our Customer Care Department within the U.S. at (800) 762-2974, outside the U.S. at (317) 572-3993 or fax (317) 572-4002.

### ***Library of Congress Cataloging-in-Publication Data:***

The data warehouse lifecycle toolkit / Ralph Kimball... [et al.]. -- 2nd ed.

p. cm.

Includes index.

ISBN 978-0-470-14977-5 (paper/website)

1. Data warehousing. I. Kimball, Ralph.

QA76.9.D37D38 2007

005.74--dc22

2007040691

**Trademarks:** Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

---

## About the Authors

---

The authors' professional careers have followed remarkably similar paths. Each author has focused on data warehousing and business intelligence (DW/BI) consulting and education for more than fifteen years. Most worked together at Metaphor Computer Systems, a pioneering decision support vendor, in the 1980s. All the authors are members of the Kimball Group and teach for Kimball University. They contribute regularly to *Intelligent Enterprise* magazine and other industry publications; most have previously written books in the *Toolkit* series.

**Ralph Kimball** founded the Kimball Group. Since the mid 1980s, he has been the DW/BI industry's thought leader on the dimensional approach and trained more than 10,000 IT professionals. Ralph has his Ph.D. in Electrical Engineering from Stanford University.

**Margy Ross** is President of the Kimball Group. She has focused exclusively on DW/BI since 1982 with an emphasis on business requirements analysis and dimensional modeling. Margy graduated with a BS in Industrial Engineering from Northwestern University.

**Warren Thornthwaite** began his DW/BI career in 1980. After managing Metaphor's consulting organization, he worked for Stanford University and WebTV. Warren holds a BA in Communications Studies from the University of Michigan and an MBA from the University of Pennsylvania's Wharton School.

**Joy Mundy** has focused on DW/BI systems since 1992 with stints at Stanford, Web TV, and Microsoft's SQL Server product development organization. Joy graduated from Tufts University with a BA in Economics, and from Stanford University with an MS in Engineering Economic Systems.

**Bob Becker** has helped clients across a variety of industries with their DW/BI challenges and solutions since 1989, including extensive work with health care organizations. Bob has a BSB in Marketing from the University of Minnesota's School of Business.





---

# Credits

---

**Executive Editor**

Robert Elliott

**Development Editor**

Sara Shlaer

**Production Editor**

Debra Banninger

**Copy Editor**

Kim Cofer

**Editorial Manager**

Mary Beth Wakefield

**Production Manager**

Tim Tate

**Vice President and Executive**

**Group Publisher**

Richard Swadley

**Vice President and Executive  
Publisher**

Joseph B. Wikert

**Project Coordinator, Cover**

Lynsey Osborn

**Proofreader**

Nancy Carrasco

**Indexer**

Melanie Belkin

**Anniversary Logo Design**

Richard Pacifico

**Cover Image**

© Steve Allen/Getty Images



---

# Acknowledgments

---

First, thanks to our students, clients, readers, and former colleagues for supporting, teaching, and influencing us. One of the authors recently received a fortune cookie that read, “You learn most when teaching others.” We couldn’t agree more. Our Kimball University students have pushed us to provide precise, specific guidance and kept us on our toes with their questions. Similarly, the challenges faced by our Kimball Group consulting clients have become our challenges, and have kept us grounded in reality. Finally, ex-colleagues have contributed to our thinking about the concepts in this book, including Laura Reeves who participated as a co-author of the first edition of the *Lifecycle Toolkit*. Beginning with our associates from the early days at Metaphor, through Red Brick, Stanford University, DecisionWorks Consulting, InfoDynamics, and Microsoft, we’ve learned lots from each of you.

Thanks to the Wiley team for making this book a reality. Bob Elliott’s subtle, yet persistent prodding got the project off the ground. Sara Shlaer did a wonderful job editing our text with an incredible amount of patience, tenacity, and attention to detail. Deb Banninger and the behind-the-scenes folks worked tirelessly to deliver a quality product. We’ve enjoyed working with all of you.

Finally, thanks to our spouses, partners, and children for putting up with the demands of our careers, while supporting us unconditionally. You’ve suffered through late nights and missed vacations alongside us. Thanks to Julie Kimball, Sara Kimball Smith, and Brian Kimball, Scott and Katie Ross, Elizabeth Wright, Tony Navarrete, and Pam, Elisa, and Jenna Becker. We couldn’t have done it without you!



---

# Contents at a Glance

---

|                   |                                                                     |            |
|-------------------|---------------------------------------------------------------------|------------|
| <b>Chapter 1</b>  | <b>Introducing the Kimball Lifecycle</b>                            | <b>1</b>   |
| <b>Chapter 2</b>  | <b>Launching and Managing the Project/Program</b>                   | <b>15</b>  |
| <b>Chapter 3</b>  | <b>Collecting the Requirements</b>                                  | <b>63</b>  |
| <b>Chapter 4</b>  | <b>Introducing the Technical Architecture</b>                       | <b>109</b> |
| <b>Chapter 5</b>  | <b>Creating the Architecture Plan and Selecting Products</b>        | <b>179</b> |
| <b>Chapter 6</b>  | <b>Introducing Dimensional Modeling</b>                             | <b>233</b> |
| <b>Chapter 7</b>  | <b>Designing the Dimensional Model</b>                              | <b>287</b> |
| <b>Chapter 8</b>  | <b>Designing the Physical Database and Planning for Performance</b> | <b>327</b> |
| <b>Chapter 9</b>  | <b>Introducing Extract, Transformation, and Load</b>                | <b>369</b> |
| <b>Chapter 10</b> | <b>Designing and Developing the ETL System</b>                      | <b>425</b> |
| <b>Chapter 11</b> | <b>Introducing Business Intelligence Applications</b>               | <b>473</b> |
| <b>Chapter 12</b> | <b>Designing and Developing Business Intelligence Applications</b>  | <b>505</b> |
| <b>Chapter 13</b> | <b>Deploying and Supporting the DW/BI System</b>                    | <b>541</b> |
| <b>Chapter 14</b> | <b>Expanding the DW/BI System</b>                                   | <b>579</b> |



---

# Contents

---

|                                                    |             |
|----------------------------------------------------|-------------|
| <b>Acknowledgments</b>                             | <b>ix</b>   |
| <b>Introduction</b>                                | <b>xxxi</b> |
| <b>Chapter 1 Introducing the Kimball Lifecycle</b> | <b>1</b>    |
| Lifecycle History Lesson                           | 1           |
| Lifecycle Milestones                               | 3           |
| Program/Project Planning                           | 4           |
| Program/Project Management                         | 4           |
| Business Requirements Definition                   | 5           |
| Technology Track                                   | 5           |
| Technical Architecture Design                      | 5           |
| Product Selection and Installation                 | 6           |
| Data Track                                         | 6           |
| Dimensional Modeling                               | 6           |
| Physical Design                                    | 6           |
| ETL Design and Development                         | 7           |
| Business Intelligence Application Track            | 7           |
| BI Application Design                              | 7           |
| BI Application Development                         | 7           |
| Deployment                                         | 7           |
| Maintenance                                        | 8           |
| Growth                                             | 8           |
| Using the Lifecycle Roadmap                        | 8           |
| Lifecycle Navigation Aids                          | 9           |
| Lifecycle Vocabulary Primer                        | 9           |
| Data Warehouse versus Business Intelligence        | 10          |
| ETL System                                         | 11          |
| Business Process Dimensional Model                 | 12          |
| Business Intelligence Applications                 | 13          |
| Conclusion                                         | 14          |

|                                                       |                                                   |           |
|-------------------------------------------------------|---------------------------------------------------|-----------|
| <b>Chapter 2</b>                                      | <b>Launching and Managing the Project/Program</b> | <b>15</b> |
| Define the Project                                    |                                                   | 16        |
| Assess Your Readiness for DW/BI                       |                                                   | 16        |
| Strong Senior Business Management Sponsor(s)          |                                                   | 16        |
| Compelling Business Motivation                        |                                                   | 17        |
| Feasibility                                           |                                                   | 17        |
| Factors Not Considered Readiness Deal Breakers        |                                                   | 18        |
| Address Shortfalls and Determine Next Steps           |                                                   | 18        |
| Strong Sponsor, Compelling Business Need, and Quality |                                                   |           |
| Data                                                  |                                                   | 19        |
| Poor Quality Data                                     |                                                   | 19        |
| Weak Business Sponsor or IT-Only Sponsor              |                                                   | 19        |
| Too Much Demand from Multiple Business Sponsors       |                                                   | 20        |
| Well Meaning, But Overly Aggressive Business Sponsor  |                                                   | 21        |
| Legacy of Underperforming, Isolated Data Silos        |                                                   | 21        |
| Develop the Preliminary Scope and Charter             |                                                   | 22        |
| Focus on a Single Business Process                    |                                                   | 22        |
| The Role of Rapid Application Development             |                                                   | 24        |
| Document the Scope/Charter                            |                                                   | 25        |
| Build the Business Case and Justification             |                                                   | 27        |
| Determine the Financial Investments and Costs         |                                                   | 27        |
| Determine the Financial Returns and Benefits          |                                                   | 28        |
| Combine the Investments and Returns to Calculate ROI  |                                                   | 30        |
| Plan the Project                                      |                                                   | 31        |
| Establish the Project Identity                        |                                                   | 31        |
| Staff the Project                                     |                                                   | 32        |
| Front Office: Sponsors and Drivers                    |                                                   | 33        |
| Coaches: Project Managers and Leads                   |                                                   | 34        |
| Regular Lineup: Core Project Team                     |                                                   | 35        |
| Special Teams                                         |                                                   | 38        |
| Free Agents                                           |                                                   | 39        |
| Convert Individual Talent into a Team                 |                                                   | 40        |
| Develop the Project Plan                              |                                                   | 40        |
| Develop the Communication Plan                        |                                                   | 43        |
| Project Team                                          |                                                   | 44        |
| Sponsor and Driver Briefings                          |                                                   | 45        |
| Business User Community                               |                                                   | 45        |
| Communication with Other Interested Parties           |                                                   | 46        |
| Manage the Project                                    |                                                   | 46        |
| Conduct the Project Team Kickoff Meeting              |                                                   | 47        |
| Monitor Project Status                                |                                                   | 48        |
| Project Status Meetings                               |                                                   | 48        |
| Project Status Reports                                |                                                   | 49        |
| Maintain the Project Plan                             |                                                   | 50        |
| Consolidate the Project Documentation                 |                                                   | 50        |



|                                                   |           |
|---------------------------------------------------|-----------|
| Manage the Scope                                  | 50        |
| Track Issues                                      | 51        |
| Control Changes                                   | 52        |
| Manage Expectations                               | 53        |
| Recognize Project Trouble Signs                   | 53        |
| Manage the Program                                | 54        |
| Establish Governance Responsibility and Processes | 54        |
| Elevate Data Stewardship to Enterprise Level      | 56        |
| Leverage Methods and Architectural Best Practices | 57        |
| Conduct Periodic Assessments                      | 57        |
| Communicate, Communicate, Communicate             | 58        |
| Conclusion                                        | 58        |
| Managing the Effort and Reducing Risk             | 58        |
| Assuring Quality                                  | 59        |
| Key Roles                                         | 59        |
| Key Deliverables                                  | 59        |
| Estimating Considerations                         | 60        |
| Website Resources                                 | 60        |
| Task List                                         | 61        |
| <b>Chapter 3 Collecting the Requirements</b>      | <b>63</b> |
| Overall Approach to Requirements Definition       | 64        |
| Interviews versus Facilitated Sessions            | 66        |
| Methods to Avoid for Collecting Requirements      | 67        |
| Prepare for the Interview                         | 68        |
| Identify the Interview Team                       | 68        |
| Lead Interviewer                                  | 68        |
| Scribe                                            | 68        |
| Observers                                         | 70        |
| Research the Organization                         | 70        |
| Select the Interviewees                           | 71        |
| Business Interviewees                             | 71        |
| IT and Compliance/Security Interviewees           | 72        |
| Develop the Interview Questionnaires              | 73        |
| Schedule the Interviews                           | 73        |
| Sequence the Interviews                           | 73        |
| Establish the Interview Time and Place            | 75        |
| Prepare the Interviewees                          | 76        |
| Review Interviewing Ground Rules                  | 77        |
| Remember Your Interview Role                      | 78        |
| Assume You Will Learn                             | 78        |
| Verify Communications                             | 78        |
| Be Conversational                                 | 79        |
| Maintain Interview Schedule Flexibility           | 79        |
| Manage Expectations Continuously                  | 80        |
| Conduct the Interview                             | 80        |
| Program Business Interviews                       | 82        |

|                                                            |            |
|------------------------------------------------------------|------------|
| Program IT Interviews                                      | 83         |
| Program Compliance/Security Interviews                     | 83         |
| Wrap Up the Interview                                      | 84         |
| Determine the Success Criteria                             | 84         |
| Say Thanks and See You Later                               | 85         |
| Review the Interview Results                               | 85         |
| Synthesize Around Business Processes                       | 86         |
| Prepare and Publish Requirements Deliverables              | 87         |
| Interview Write-Ups                                        | 88         |
| Program Requirements Findings Document                     | 88         |
| Prioritize and Agree on Next Steps                         | 91         |
| Culminate with a Review and Prioritization Meeting         | 91         |
| Close the Loop                                             | 93         |
| Adjustments for Project Level Requirements                 | 93         |
| Project Level Approach                                     | 93         |
| Prepare for the Project Requirements Interview             | 94         |
| Select the Interviewees                                    | 95         |
| Prepare the Interviewees                                   | 95         |
| Conduct the Interviews                                     | 97         |
| Review Existing Reports and Analyses                       | 98         |
| Wrap Up the Interview                                      | 99         |
| Dig into the Data                                          | 99         |
| Review the Interview Results                               | 100        |
| Prepare and Publish Project Deliverables                   | 100        |
| Agree on Next Steps and Close the Loop                     | 101        |
| Deal with Challenging Interviewees                         | 101        |
| Abused User                                                | 101        |
| Overbooked/Substitute User                                 | 102        |
| Comatose User                                              | 102        |
| Overzealous User                                           | 102        |
| Know-It-All User                                           | 103        |
| Clueless User                                              | 103        |
| Nonexistent User                                           | 103        |
| Conclusion                                                 | 104        |
| Managing the Effort and Reducing Risk                      | 104        |
| Assuring Quality                                           | 105        |
| Key Roles                                                  | 105        |
| Key Deliverables                                           | 105        |
| Estimating Considerations                                  | 106        |
| Website Resources                                          | 106        |
| Task List                                                  | 107        |
| <b>Chapter 4    Introducing the Technical Architecture</b> | <b>109</b> |
| The Value of Architecture                                  | 110        |
| Technical Architecture Overview                            | 112        |
| Flow from Source System to User Screen                     | 113        |
| Common Architecture Features                               | 115        |

|                                                   |     |
|---------------------------------------------------|-----|
| Metadata Driven                                   | 115 |
| Flexible Services Layers                          | 117 |
| Evolution of Your DW/BI Architecture              | 118 |
| Back Room Architecture                            | 119 |
| General ETL Requirements                          | 119 |
| Build versus Buy                                  | 120 |
| Back Room ETL Flow                                | 121 |
| Source Systems                                    | 121 |
| Enterprise Resource Planning Systems              | 123 |
| Operational Data Stores                           | 124 |
| Reporting Operational Data Stores                 | 125 |
| Master Data Management                            | 125 |
| XML Sources                                       | 126 |
| Message Queues, Log Files, and Redo Files         | 126 |
| Proprietary Formats                               | 126 |
| Extract                                           | 127 |
| Clean and Conform                                 | 127 |
| Deliver                                           | 127 |
| ETL Management Services                           | 128 |
| Additional Back Room Services and Trends          | 129 |
| Data Service Providers                            | 129 |
| Functional Service Providers                      | 129 |
| Data Delivery Services                            | 129 |
| ETL Data Stores                                   | 130 |
| ETL System Data Stores                            | 130 |
| Lookup and Decode Tables                          | 130 |
| Data Quality Data Stores                          | 131 |
| ETL Metadata                                      | 131 |
| Back Room Summary                                 | 132 |
| Presentation Server Architecture                  | 133 |
| Business Requirements for Information             | 133 |
| Detail Atomic Data                                | 134 |
| Aggregates                                        | 134 |
| Aggregate Navigation                              | 136 |
| Design Disciplines within the Presentation Server | 138 |
| Adjusting the Presentation Server Architecture    | 138 |
| Organizational Considerations                     | 139 |
| Presentation Server Metadata                      | 140 |
| Presentation Server Summary                       | 141 |
| Front Room Architecture                           | 141 |
| BI Application Types                              | 141 |
| BI Management Services                            | 143 |
| Shared Services                                   | 143 |
| Vendor Specific Architectural Choices             | 150 |
| BI Data Stores                                    | 151 |
| Stored Reports                                    | 151 |

|                                                          |     |
|----------------------------------------------------------|-----|
| Application Server Caches                                | 151 |
| Local User Databases                                     | 151 |
| Disposable Analytic Data Stores                          | 152 |
| Results from Analytic Applications                       | 152 |
| Downstream Systems                                       | 153 |
| Data Store Security                                      | 153 |
| Desktop Tool Architecture Approaches                     | 154 |
| BI Metadata                                              | 154 |
| Front Room Summary                                       | 155 |
| Infrastructure                                           | 156 |
| Infrastructure Drivers                                   | 156 |
| Back Room and Presentation Server Infrastructure Factors | 157 |
| Parallel Processing Hardware Architectures               | 159 |
| Symmetric Multiprocessing (SMP)                          | 159 |
| Massively Parallel Processing (MPP)                      | 161 |
| Non-Uniform Memory Architecture (NUMA)                   | 161 |
| Clusters                                                 | 161 |
| Warehouse Appliances                                     | 162 |
| Partitioning Hardware                                    | 163 |
| Considerations Common to All Parallel Architectures      | 163 |
| Hardware Performance Boosters                            | 163 |
| Disk Issues                                              | 163 |
| Memory                                                   | 164 |
| CPUs                                                     | 165 |
| Secondary Storage                                        | 165 |
| Database Platform Factors                                | 165 |
| Characteristics of Relational Engines                    | 165 |
| Characteristics of OLAP Engines                          | 166 |
| Front Room Infrastructure Factors                        | 167 |
| Application Server Considerations                        | 167 |
| Desktop Considerations                                   | 168 |
| Connectivity and Networking Factors                      | 169 |
| Infrastructure Summary                                   | 170 |
| Metadata                                                 | 170 |
| Value of Metadata Integration                            | 171 |
| Impact Analysis                                          | 171 |
| Audit and Documentation                                  | 171 |
| Metadata Quality and Management                          | 171 |
| Options for Metadata Integration                         | 172 |
| Single Source DW/BI System Vendors                       | 172 |
| Core Vendor Product                                      | 172 |
| Do It Yourself                                           | 173 |
| Metadata Summary                                         | 173 |
| Security                                                 | 173 |
| Security Vulnerabilities                                 | 174 |
| Threats to Physical Assets                               | 174 |

|                                                                           |            |
|---------------------------------------------------------------------------|------------|
| Threats to Information and Software Assets                                | 175        |
| Threats to Business Continuity                                            | 176        |
| Network Threats                                                           | 176        |
| Security Summary                                                          | 177        |
| Conclusion                                                                | 178        |
| <b>Chapter 5    Creating the Architecture Plan and Selecting Products</b> | <b>179</b> |
| Create the Architecture                                                   | 180        |
| Architecture Development Process                                          | 180        |
| Develop the Application Architecture Plan                                 | 183        |
| Step 1 — Form an Architecture Task Force                                  | 183        |
| Step 2 — Gather Architecture-Related Requirements                         | 184        |
| Step 3 — Create a Draft Architectural Implications Document               | 184        |
| Step 4 — Create the Architecture Model                                    | 185        |
| Step 5 — Determine the Architecture Implementation Phases                 | 186        |
| Step 6 — Design and Specify the Subsystems                                | 187        |
| Step 7 — Create the Application Architecture Plan Document                | 187        |
| Step 8 — Review the Draft                                                 | 187        |
| Example Application Architecture Plan Outline and Model                   | 188        |
| Select Products                                                           | 191        |
| Keep a Business Focus                                                     | 191        |
| Major DW/BI Evaluation Areas                                              | 192        |
| Evaluate Options and Choose a Product                                     | 192        |
| Step 1 — Understand the Purchasing Process                                | 193        |
| Step 2 — Develop the Product Evaluation Matrix                            | 193        |
| Step 3 — Conduct Market Research                                          | 194        |
| Step 4 — Narrow Your Options to a Short List                              | 196        |
| Step 5 — Evaluate the Candidates                                          | 197        |
| Step 6 — Recommend a Product                                              | 201        |
| Step 7 — Trial                                                            | 201        |
| Step 8 — Contract Negotiations                                            | 201        |
| Considerations for the Back Room and Presentation Server                  | 202        |
| Hardware Platform                                                         | 202        |
| DBMS Platform                                                             | 203        |
| ETL Tool                                                                  | 204        |
| Considerations for the Front Room                                         | 205        |
| Manage the Metadata                                                       | 207        |
| Appoint the Metadata Manager                                              | 207        |
| Create the Metadata Strategy                                              | 208        |
| Secure the System                                                         | 210        |
| Secure the Hardware and Operating System                                  | 211        |
| Secure the Development Environment                                        | 211        |
| Secure the Network                                                        | 211        |
| Network Components                                                        | 212        |

|                                                   |            |
|---------------------------------------------------|------------|
| Encryption                                        | 214        |
| Authenticate the Users                            | 215        |
| Secure the Data                                   | 215        |
| Provide Open Access for Internal Users            | 215        |
| Itemize Sensitive Data                            | 216        |
| Minimize or Mask Sensitive Data                   | 216        |
| Secure the Data Access                            | 217        |
| Monitor Usage and Ensure Compliance               | 220        |
| Plan for Backup and Recovery                      | 220        |
| Create the Infrastructure Map                     | 221        |
| Install the Hardware and Software                 | 224        |
| Conclusion                                        | 226        |
| Managing the Effort and Reducing Risk             | 226        |
| Assuring Quality                                  | 227        |
| Key Roles                                         | 227        |
| Key Deliverables                                  | 228        |
| Estimating Considerations                         | 228        |
| Creating the Architecture Plan                    | 228        |
| Selecting Products                                | 229        |
| Metadata                                          | 229        |
| Security                                          | 229        |
| Website Resources                                 | 229        |
| Task List                                         | 230        |
| <b>Chapter 6 Introducing Dimensional Modeling</b> | <b>233</b> |
| Making the Case for Dimensional Modeling          | 234        |
| What Is Dimensional Modeling?                     | 234        |
| What about Normalized Modeling?                   | 235        |
| Benefits of Dimensional Modeling                  | 237        |
| Dimensional Modeling Primer                       | 238        |
| Fact Tables                                       | 238        |
| Fact Table Keys                                   | 240        |
| Fact Table Granularity                            | 240        |
| Dimension Tables                                  | 241        |
| Dimension Table Keys                              | 243        |
| Conformed Dimensions                              | 244        |
| Four-Step Dimensional Design Process              | 246        |
| Step 1 — Choose the Business Process              | 246        |
| Step 2 — Declare the Grain                        | 246        |
| Step 3 — Identify the Dimensions                  | 247        |
| Step 4 — Identify the Facts                       | 247        |
| Enterprise Data Warehouse Bus Architecture        | 248        |
| Planning Crisis                                   | 248        |
| Bus Architecture                                  | 249        |
| Value Chain Implications                          | 250        |
| Common Matrix Mishaps                             | 252        |
| Taking the Pledge                                 | 253        |

|                                                     |            |
|-----------------------------------------------------|------------|
| More on Dimensions                                  | 253        |
| Date and Time                                       | 253        |
| Surrogate Date Keys                                 | 254        |
| Time of Day                                         | 255        |
| Date/Timestamps                                     | 255        |
| Multiple Time Zones                                 | 256        |
| Degenerate Dimensions                               | 256        |
| Slowly Changing Dimensions                          | 257        |
| Type 1: Overwrite the Dimension Attribute           | 257        |
| Type 2: Add a New Dimension Row                     | 258        |
| Type 3: Add a New Dimension Attribute               | 258        |
| Mini-Dimensions: Add a New Dimension                | 259        |
| Hybrid Slowly Changing Dimension Techniques         | 261        |
| Role-Playing Dimensions                             | 262        |
| Junk Dimensions                                     | 263        |
| Snowflaking and Outriggers                          | 265        |
| Handling Hierarchies                                | 268        |
| Fixed Hierarchies                                   | 268        |
| Variable Depth Hierarchies via Bridge Tables        | 268        |
| Many-Valued Dimensions with Bridge Tables           | 270        |
| More on Facts                                       | 273        |
| Three Fundamental Grains                            | 273        |
| Transaction Fact Tables                             | 273        |
| Periodic Snapshot Fact Tables                       | 274        |
| Accumulating Snapshot Fact Tables                   | 274        |
| Facts of Differing Granularity and Allocation       | 276        |
| Multiple Currencies and Units of Measure            | 278        |
| Factless Fact Tables                                | 280        |
| Consolidated Fact Tables                            | 281        |
| Fables and Falsehoods About Dimensional Modeling    | 282        |
| Fables Caused by Focusing on Departmental Reports   | 282        |
| Fables Caused by Premature Summarization            | 284        |
| Fables Caused by Overvaluing Normalization          | 285        |
| Conclusion                                          | 286        |
| <b>Chapter 7    Designing the Dimensional Model</b> | <b>287</b> |
| Modeling Process Overview                           | 288        |
| Get Organized                                       | 289        |
| Identify Design Participants                        | 289        |
| Revisit the Requirements                            | 292        |
| Use Modeling Tools                                  | 293        |
| Establish Naming Conventions                        | 295        |
| Provide for Source Data Research and Profiling      | 296        |
| Obtain Facilities and Supplies                      | 297        |
| Recall the Four-Step Modeling Process               | 297        |
| Step 1 — Choose the Business Process                | 298        |
| Step 2 — Declare the Grain                          | 298        |

|                                                                         |            |
|-------------------------------------------------------------------------|------------|
| Step 3 — Identify the Dimensions                                        | 299        |
| Step 4 — Identify the Facts                                             | 301        |
| Design the Dimensional Model                                            | 301        |
| Build the High Level Dimensional Model                                  | 302        |
| Conduct the Initial Design Session                                      | 302        |
| Document the High Level Model Diagram                                   | 303        |
| Identify the Attributes and Metrics                                     | 304        |
| Develop the Detailed Dimensional Model                                  | 305        |
| Identify the Data Sources                                               | 305        |
| Establish Conformed Dimensions                                          | 309        |
| Identify Base Facts and Derived Facts                                   | 310        |
| Document the Detailed Table Designs                                     | 311        |
| Update the Bus Matrix                                                   | 314        |
| Identify and Resolve Modeling Issues                                    | 314        |
| Review and Validate the Model                                           | 316        |
| Perform IT Data Model Review                                            | 318        |
| Review with Core Users                                                  | 319        |
| Present to the Business Users                                           | 319        |
| Finalize the Design Documentation                                       | 320        |
| Embrace Data Stewardship                                                | 321        |
| Conclusion                                                              | 322        |
| Managing the Effort and Reducing Risk                                   | 323        |
| Assuring Quality                                                        | 323        |
| Key Roles                                                               | 323        |
| Key Deliverables                                                        | 324        |
| Estimating Considerations                                               | 324        |
| Website Resources                                                       | 324        |
| Task List                                                               | 325        |
| <br><b>Chapter 8</b>                                                    |            |
| <b>    Designing the Physical Database and Planning for Performance</b> | <b>327</b> |
| Develop Standards                                                       | 328        |
| Follow Naming Conventions                                               | 329        |
| To Null or Not to Null?                                                 | 330        |
| Place Staging Tables                                                    | 330        |
| Develop File Location Standards                                         | 331        |
| Use Synonyms or Views for User Accessible Tables                        | 331        |
| Primary Keys                                                            | 332        |
| Foreign Keys                                                            | 334        |
| Develop the Physical Data Model                                         | 335        |
| Design the Physical Data Structure                                      | 335        |
| Finalize the Source-to-Target Map                                       | 336        |
| Star versus Snowflake                                                   | 338        |
| Use a Data Modeling Tool                                                | 340        |
| Develop Initial Sizing Estimates                                        | 340        |
| Build the Development Database                                          | 343        |
| Design Processing Data Stores                                           | 343        |



|                                                                |            |
|----------------------------------------------------------------|------------|
| Develop the Initial Index Plan                                 | 344        |
| Indexing and Query Strategy Overview                           | 345        |
| B-Tree Index                                                   | 345        |
| Clustered Index                                                | 345        |
| Bitmapped Index                                                | 346        |
| Other Index Types                                              | 346        |
| Star Schema Optimization                                       | 347        |
| Indexing Dimension Tables                                      | 347        |
| Indexing Fact Tables                                           | 348        |
| Indexing for Loads                                             | 349        |
| Indexing for OLAP                                              | 349        |
| Analyze Tables and Indexes after the Load                      | 349        |
| Design the OLAP Database                                       | 350        |
| OLAP Data Granularity and Drillthrough                         | 350        |
| Perfecting the OLAP Dimensions                                 | 351        |
| Defining OLAP Calculations                                     | 352        |
| Build the Test Database                                        | 352        |
| Design Aggregations                                            | 353        |
| Deciding How to Aggregate                                      | 353        |
| Deciding What to Aggregate                                     | 354        |
| Maintaining Aggregations                                       | 355        |
| Finalizing Indexes                                             | 357        |
| Design and Build the Database Instance                         | 357        |
| Memory                                                         | 358        |
| Block Size                                                     | 358        |
| Save the Database Build Scripts and Parameter Files            | 358        |
| Develop the Physical Storage Structure                         | 359        |
| Compute Table and Index Sizes                                  | 359        |
| Develop the Partitioning Plan                                  | 360        |
| Set up Storage                                                 | 361        |
| Fault Tolerance                                                | 362        |
| Storage Area Networks                                          | 362        |
| Configuration of Volumes and Drives                            | 362        |
| Conclusion                                                     | 363        |
| Managing the Effort and Reducing Risk                          | 364        |
| Assuring Quality                                               | 364        |
| Key Roles                                                      | 365        |
| Key Deliverables                                               | 365        |
| Estimating Considerations                                      | 365        |
| Website Resources                                              | 366        |
| Task List                                                      | 367        |
| <b>Chapter 9 Introducing Extract, Transformation, and Load</b> | <b>369</b> |
| Round Up the Requirements                                      | 370        |
| Business Needs                                                 | 370        |
| Compliance                                                     | 371        |
| Data Quality                                                   | 371        |

|                                                               |     |
|---------------------------------------------------------------|-----|
| Security                                                      | 372 |
| Data Integration                                              | 372 |
| Data Latency                                                  | 373 |
| Archiving and Lineage                                         | 373 |
| User Delivery Interfaces                                      | 373 |
| Available Skills                                              | 374 |
| Legacy Licenses                                               | 374 |
| The 34 Subsystems of ETL                                      | 374 |
| Extracting Data                                               | 375 |
| Subsystem 1 — Data Profiling                                  | 375 |
| Subsystem 2 — Change Data Capture System                      | 376 |
| Subsystem 3 — Extract System                                  | 378 |
| Cleaning and Conforming Data                                  | 380 |
| Improving Your Data Quality Culture and Processes             | 380 |
| Subsystem 4 — Data Cleansing System                           | 381 |
| Quality Screens                                               | 382 |
| Responding to Quality Events                                  | 383 |
| Subsystem 5 — Error Event Schema                              | 383 |
| Subsystem 6 — Audit Dimension Assembler                       | 385 |
| Subsystem 7 — Deduplication System                            | 386 |
| Subsystem 8 — Conforming System                               | 386 |
| Delivering Data for Presentation                              | 387 |
| Subsystem 9 — Slowly Changing Dimension Manager               | 387 |
| Type 1: Overwrite                                             | 389 |
| Type 2: Create a New Row                                      | 390 |
| Type 3: Add a New Column                                      | 391 |
| Hybrid: Combination of Types                                  | 391 |
| Subsystem 10 — Surrogate Key Generator                        | 392 |
| Subsystem 11 — Hierarchy Manager                              | 393 |
| Subsystem 12 — Special Dimensions Manager                     | 393 |
| Subsystem 13 — Fact Table Builders                            | 395 |
| Transaction Grain Fact Table Loader                           | 396 |
| Periodic Snapshot Fact Table Loader                           | 396 |
| Accumulating Snapshot Fact Table Loader                       | 397 |
| Subsystem 14 — Surrogate Key Pipeline                         | 398 |
| Subsystem 15 — Multi-Valued Dimension Bridge Table<br>Builder | 400 |
| Subsystem 16 — Late Arriving Data Handler                     | 400 |
| Subsystem 17 — Dimension Manager System                       | 402 |
| Subsystem 18 — Fact Provider System                           | 402 |
| Subsystem 19 — Aggregate Builder                              | 403 |
| Subsystem 20 — OLAP Cube Builder                              | 404 |
| Subsystem 21 — Data Propagation Manager                       | 404 |
| Managing the ETL Environment                                  | 405 |
| Subsystem 22 — Job Scheduler                                  | 406 |
| Subsystem 23 — Backup System                                  | 407 |

|                                                           |            |
|-----------------------------------------------------------|------------|
| Backup                                                    | 407        |
| Archive and Retrieval                                     | 408        |
| Subsystem 24 — Recovery and Restart System                | 409        |
| Subsystem 25 — Version Control System                     | 410        |
| Subsystem 26 — Version Migration System                   | 410        |
| Subsystem 27 — Workflow Monitor                           | 411        |
| Subsystem 28 — Sorting System                             | 412        |
| Subsystem 29 — Lineage and Dependency Analyzer            | 412        |
| Subsystem 30 — Problem Escalation System                  | 413        |
| Subsystem 31 — Parallelizing/Pipelining System            | 414        |
| Subsystem 32 — Security System                            | 415        |
| Subsystem 33 — Compliance Manager                         | 415        |
| Subsystem 34 — Metadata Repository Manager                | 417        |
| Real Time Implications                                    | 417        |
| Real Time Triage                                          | 417        |
| Real Time Tradeoffs                                       | 419        |
| Real Time Partitions in the Presentation Server           | 421        |
| Transaction Grain Real Time Partition                     | 421        |
| Periodic Snapshot Real Time Partition                     | 422        |
| Conclusion                                                | 423        |
| <b>Chapter 10 Designing and Developing the ETL System</b> | <b>425</b> |
| ETL Process Overview                                      | 425        |
| Getting Started                                           | 426        |
| Develop the ETL Plan                                      | 428        |
| Step 1 — Draw the High Level Plan                         | 428        |
| Step 2 — Choose an ETL Tool                               | 429        |
| Step 3 — Develop Default Strategies                       | 430        |
| Step 4 — Drill Down by Target Table                       | 432        |
| Ensure Clean Hierarchies                                  | 432        |
| Develop Detailed Table Schematics                         | 434        |
| Develop the ETL Specification Document                    | 435        |
| Develop a Sandbox Source System                           | 436        |
| Develop One-Time Historic Load Processing                 | 437        |
| Step 5 — Populate Dimension Tables with Historic Data     | 438        |
| Populate Type 1 Dimension Tables                          | 438        |
| Dimension Transformations                                 | 439        |
| Dimension Table Loading                                   | 443        |
| Load Type 2 Dimension Table History                       | 445        |
| Populate Date and Other Static Dimensions                 | 446        |
| Step 6 — Perform the Fact Table Historic Load             | 447        |
| Historic Fact Table Extracts                              | 447        |
| Fact Table Transformations                                | 448        |
| Fact Table Loading                                        | 454        |
| Test, Test, and Test Again                                | 455        |
| Develop Incremental ETL Processing                        | 456        |
| Step 7 — Dimension Table Incremental Processing           | 456        |

|                                                                  |            |
|------------------------------------------------------------------|------------|
| Dimension Table Extracts                                         | 456        |
| Identify New and Changed Dimension Rows                          | 457        |
| Process Changes to Dimension Attributes                          | 458        |
| Step 8 — Fact Table Incremental Processing                       | 459        |
| Fact Table Extract and Data Quality Checkpoint                   | 460        |
| Fact Table Transformations and Surrogate Key Pipeline            | 461        |
| Late Arriving Facts and the Surrogate Key Pipeline               | 462        |
| Incremental Fact Table Load                                      | 463        |
| Speed Up the Load Cycle                                          | 464        |
| Step 9 — Aggregate Table and OLAP Loads                          | 465        |
| Step 10 — ETL System Operation and Automation                    | 466        |
| Schedule Jobs                                                    | 466        |
| Handle Predictable Exceptions and Errors Automatically           | 467        |
| Handle Unpredictable Errors Gracefully                           | 467        |
| Maintain Database Objects                                        | 468        |
| Develop and Test ETL Automation                                  | 468        |
| Conclusion                                                       | 469        |
| Managing the Effort and Reducing Risk                            | 469        |
| Assuring Quality                                                 | 470        |
| Key Roles                                                        | 470        |
| Key Deliverables                                                 | 471        |
| Estimating Considerations                                        | 471        |
| Website Resources                                                | 471        |
| Task List                                                        | 472        |
| <b>Chapter 11 Introducing Business Intelligence Applications</b> | <b>473</b> |
| Importance of Business Intelligence Applications                 | 474        |
| Analytic Cycle for Business Intelligence                         | 476        |
| Stage 1: Monitor Activity                                        | 476        |
| Stage 2: Identify Exceptions                                     | 477        |
| Stage 3: Determine Causal Factors                                | 477        |
| Stage 4: Model Alternatives                                      | 478        |
| Stage 5: Take Action and Track Results                           | 478        |
| More Implications of the Analytic Cycle                          | 479        |
| Types of Business Intelligence Applications                      | 479        |
| Direct Access Query and Reporting Tools                          | 480        |
| Query Formulation                                                | 481        |
| Analysis and Presentation Capabilities                           | 483        |
| User Experience                                                  | 485        |
| Technical Features                                               | 486        |
| Standard Reports                                                 | 487        |
| Analytic Applications                                            | 488        |
| Pre-Built Analytic Applications                                  | 489        |
| Read/Write Analytic Applications                                 | 490        |
| Dashboards and Scorecards                                        | 490        |
| Operational Business Intelligence                                | 493        |
| Data Mining                                                      | 494        |

|                                                                               |            |
|-------------------------------------------------------------------------------|------------|
| Data Mining Overview                                                          | 494        |
| Data Mining in the Applications Architecture                                  | 498        |
| Navigating Applications via the BI Portal                                     | 499        |
| Density Considerations                                                        | 501        |
| Navigation Structure Based on Business Processes                              | 502        |
| Additional Portal Functions                                                   | 502        |
| Application Interface Alternatives                                            | 503        |
| Conclusion                                                                    | 504        |
| <b>Chapter 12 Designing and Developing Business Intelligence Applications</b> | <b>505</b> |
| Business Intelligence Application Resource Planning                           | 506        |
| Role of the BI Application Developer                                          | 506        |
| Who Does the BI Applications Job?                                             | 506        |
| Lifecycle Timing                                                              | 507        |
| Business Intelligence Application Specification                               | 508        |
| Create Application Standards and Templates                                    | 508        |
| Determine Naming Standards                                                    | 508        |
| Create the Application Templates                                              | 509        |
| Create Dashboard and Analytic Application Templates                           | 512        |
| Determine the Initial Application Set                                         | 512        |
| Identify Report Candidates                                                    | 512        |
| Consolidate the Candidate List                                                | 514        |
| Prioritize the Report List                                                    | 514        |
| Develop Detailed Application Specifications                                   | 515        |
| Specify Application Content                                                   | 517        |
| Design the Navigation Framework and Portal                                    | 519        |
| Review and Validate the Applications and Model                                | 520        |
| Review with the Business                                                      | 521        |
| Business Intelligence Application Development                                 | 522        |
| Prepare for Application Development                                           | 522        |
| Install and Test the BI Tools                                                 | 522        |
| Validate Your BI Application Interface Strategy                               | 523        |
| Set Up User Security                                                          | 523        |
| Set Up the Report Process Metadata System                                     | 524        |
| Build the Applications                                                        | 524        |
| Follow the Core Process                                                       | 525        |
| Design and Develop Analytic Applications                                      | 527        |
| Design and Develop Operational BI Applications                                | 527        |
| Include Data Mining Models                                                    | 529        |
| Validate the Data and Data Model                                              | 530        |
| Create the Navigational BI Portal                                             | 530        |
| Set Up Report Scheduling                                                      | 532        |
| Test and Verify the Applications and Data                                     | 532        |
| Complete the Documentation                                                    | 534        |
| Plan for Deployment                                                           | 534        |
| Business Intelligence Application Maintenance                                 | 534        |

|                                                             |            |
|-------------------------------------------------------------|------------|
| Conclusion                                                  | 535        |
| Managing the Effort and Reducing Risk                       | 535        |
| Assuring Quality                                            | 536        |
| Key Roles                                                   | 536        |
| Key Deliverables                                            | 537        |
| Estimating Considerations                                   | 537        |
| BI Application Specification                                | 537        |
| Development                                                 | 538        |
| Website Resources                                           | 538        |
| Task List                                                   | 539        |
| <b>Chapter 13 Deploying and Supporting the DW/BI System</b> | <b>541</b> |
| System Deployment                                           | 542        |
| Pre-Deployment Testing                                      | 543        |
| System Testing Procedures                                   | 543        |
| Data Quality Assurance Testing                              | 546        |
| Operations Process Testing                                  | 547        |
| Live Testing                                                | 548        |
| Performance Testing                                         | 549        |
| Usability Testing                                           | 552        |
| Desktop Readiness and Configuration                         | 553        |
| Deployment                                                  | 554        |
| Relational Database Deployment                              | 554        |
| ETL Deployment                                              | 556        |
| OLAP Database Deployment                                    | 556        |
| Report Deployment                                           | 557        |
| Documentation and Training                                  | 558        |
| Core Documentation                                          | 558        |
| Business Process Dimensional Model Descriptions             | 558        |
| Table and Column Descriptions                               | 559        |
| Report Descriptions                                         | 559        |
| Additional Documentation                                    | 559        |
| User Training                                               | 560        |
| Design and Approach                                         | 560        |
| Develop Training Materials                                  | 561        |
| Create a Training Database                                  | 562        |
| Plan for the Level of Effort                                | 562        |
| Maintenance and Support                                     | 563        |
| Manage the Front Room                                       | 563        |
| Provide User Support                                        | 563        |
| Maintain the BI Portal                                      | 565        |
| Manage Security                                             | 566        |
| Monitor Usage                                               | 566        |
| Report on Usage                                             | 567        |
| Manage the Back Room                                        | 567        |
| Support Data Reconciliation                                 | 568        |
| Execute and Monitor the ETL System                          | 568        |