

Contents

Introduction	xxi
1 <i>The Reader</i> at a Glance	1
Setting Up for Success	1
1.1 Resist the Urge to Start Coding.....	1
1.2 Set Your Boundaries	4
Tackling DW/BI Design and Development	6
1.3 Data Wrangling.....	6
1.4 Myth Busters	8
1.5 Dividing the World	10
1.6 Essential Steps for the Integrated Enterprise Data Warehouse. . .	13
1.7 Drill Down to Ask Why	22
1.8 Slowly Changing Dimensions	24
1.9 Judge Your BI Tool through Your Dimensions	28
1.10 Fact Tables	30
1.11 Exploit Your Fact Tables.....	32
2 Before You Dive In.....	35
Historical Perspective.....	35
2.1 The Database Market Splits.....	36
2.2 Bringing Up Supermarts	38
Dealing with Demanding Realities.....	45
2.3 Brave New Requirements for Data Warehousing	46
2.4 Coping with the Brave New Requirements	50
2.5 Stirring Things Up.....	55
2.6 Design Constraints and Unavoidable Realities	58
2.7 Two Powerful Ideas.....	61
2.8 Data Warehouse Dining Experience	65

3	Project/Program Planning	69
	Professional Responsibilities	69
	3.1 Professional Boundaries	69
	3.2 An Engineer's View	72
	3.3 Beware the Objection Removers	76
	3.4 What Does the Central Team Do?	80
	3.5 Avoid DW/BI Isolation	83
	3.6 Implementation Analysis Paralysis	84
	Justification and Sponsorship	85
	3.7 Habits of Effective Sponsors	86
	3.8 TCO Starts with the End User	89
	3.9 Better Business Skills for BI and Data Warehouse Professionals	93
	Kimball Methodology	96
	3.10 Kimball Lifecycle in a Nutshell	96
	3.11 Off the Bench	99
	3.12 The Anti-Architect	100
	3.13 Think Critically When Applying Best Practices	103
	3.14 Eight Guidelines for Low Risk Enterprise Data Warehousing	106
	3.15 Relating to Agile Methodologies	109
	3.16 Is Agile Enterprise Data Warehousing an Oxymoron?	110
4	Requirements Definition	113
	Gathering Requirements	113
	4.1 Alan Alda's Interviewing Tips for Uncovering Business Requirements	113
	4.2 More Business Requirements Gathering Dos and Don'ts	117
	4.3 Overcoming Obstacles When Gathering Business Requirements	119
	4.4 Surprising Value of Data Profiling	121
	Organizing around Business Processes	123
	4.5 Focus on Business Processes, Not Business Departments!	123
	4.6 Identifying Business Processes	124
	4.7 Business Process Decoder Ring	127
	4.8 Relationship between Strategic Business Initiatives and Business Processes	127
	Wrapping Up the Requirements	128
	4.9 The Bottom-Up Misnomer	128

5	Data Architecture	133
	Making the Case for Dimensional Modeling	133
	5.1 Is ER Modeling Hazardous to DSS?	133
	5.2 A Dimensional Modeling Manifesto	137
	5.3 There Are No Guarantees	145
	Enterprise Data Warehouse Bus Architecture	148
	5.4 Divide and Conquer	148
	5.5 The Matrix.	151
	5.6 The Matrix: Revisited	155
	5.7 Drill Down into a Detailed Bus Matrix	159
	Integration Instead of Centralization.	161
	5.8 Integration for Real People	161
	5.9 Data Stewardship 101: The First Step to Quality and Consistency	165
	5.10 To Be or Not To Be Centralized	168
	Contrast with the Corporate Information Factory	171
	5.11 Differences of Opinion	171
	5.12 Don't Support Business Intelligence with a Normalized EDW	176
6	Dimensional Modeling Fundamentals	179
	Basics of Dimensional Modeling	179
	6.1 Fact Tables and Dimension Tables	179
	6.2 Drilling Down, Up, and Across	183
	6.3 The Soul of the Data Warehouse, Part One: Drilling Down	186
	6.4 The Soul of the Data Warehouse, Part Two: Drilling Across	189
	6.5 The Soul of the Data Warehouse, Part Three: Handling Time	191
	6.6 Graceful Modifications to Existing Fact and Dimension Tables	194
	Dos and Don'ts.	196
	6.7 Kimball's Ten Essential Rules of Dimensional Modeling	196
	6.8 What Not to Do.	199
	Myths about Dimensional Modeling	201
	6.9 Dangerous Preconceptions	201
	6.10 Fables and Facts	204
7	Dimensional Modeling Tasks and Responsibilities	209
	Design Activities	209
	7.1 Letting the Users Sleep	209

7.2 Staffing the Dimensional Modeling Team	216
7.3 Practical Steps for Designing a Dimensional Model	217
7.4 The Naming Game	220
7.5 When Is the Dimensional Design Done?	221
Design Review Activities	223
7.6 Fistful of Flaws	223
7.7 Rating Your Dimensional Data Warehouse	226
8 Fact Table Core Concepts	233
Granularity	233
8.1 Declaring the Grain	233
8.2 Keep to the Grain in Dimensional Modeling	236
8.3 Warning: Summary Data May Be Hazardous to Your Health	238
8.4 No Detail Too Small	239
Types of Fact Tables	242
8.5 Fundamental Grains	243
8.6 Modeling a Pipeline with an Accumulating Snapshot	246
8.7 Combining Periodic and Accumulating Snapshots	249
8.8 Modeling Time Spans	250
8.9 A Rolling Prediction of the Future, Now and in the Past	252
8.10 Factless Fact Tables	255
8.11 Factless Fact Tables? Sound Like Jumbo Shrimp?	258
8.12 What Didn't Happen	259
Parent-Child Fact Tables	262
8.13 Managing Your Parents	263
8.14 Patterns to Avoid When Modeling Header/Line Item Transactions	266
Fact Table Keys and Degenerates	268
8.15 Fact Table Surrogate Keys	268
8.16 Reader Suggestions on Fact Table Surrogate Keys	269
8.17 Another Look at Degenerate Dimensions	271
8.18 Creating a Reference Dimension for Infrequently Accessed Degenerates	272
Miscellaneous Fact Table Design Patterns	273
8.19 Put Your Fact Tables on a Diet	273

8.20 Keeping Text Out of the Fact Table 275

8.21 Dealing with Nulls in a Dimensional Model 276

8.22 Modeling Data as Both a Fact and Dimension Attribute 277

8.23 When a Fact Table Can Be Used as a Dimension Table 278

8.24 Sparse Facts and Facts with Short Lifetimes 280

8.25 Pivoting the Fact Table with a Fact Dimension 282

9 Dimension Table Core Concepts 285

Dimension Table Keys 285

 9.1 Surrogate Keys 285

 9.2 Keep Your Keys Simple 289

Date and Time Dimension Considerations 290

 9.3 It's Time for Time 291

 9.4 Surrogate Keys for the Time Dimension 293

 9.5 Latest Thinking on Time Dimension Tables 295

 9.6 Smart Date Keys to Partition Fact Tables 296

 9.7 Handling All the Dates 298

Miscellaneous Dimension Patterns 299

 9.8 Data Warehouse Role Models 299

 9.9 Mystery Dimensions 303

 9.10 De-Clutter with Junk Dimensions 306

 9.11 Showing the Correlation Between Dimensions 307

 9.12 Causal (Not Casual) Dimensions 308

 9.13 Resist Abstract Generic Dimensions 311

 9.14 Hot-Swappable Dimensions 312

 9.15 Accurate Counting with a Dimensional Supplement 314

Slowly Changing Dimensions 315

 9.16 Perfectly Partitioning History with Type 2 SCD 316

 9.17 Many Alternate Realities 316

 9.18 Monster Dimensions 320

 9.19 When a Slowly Changing Dimension Speeds Up 322

 9.20 When Do Dimensions Become Dangerous? 325

 9.21 Slowly Changing Dimensions Are Not Always as
 Easy as 1, 2, and 3 326

 9.22 Dimension Row Change Reason Attributes 330

10	More Dimension Patterns and Case Studies	333
	Snowflakes, Outriggers, and Bridges	333
	10.1 Snowflakes, Outriggers, and Bridges	333
	10.2 A Trio of Interesting Snowflakes	336
	10.3 Help for Dimensional Modeling	340
	10.4 Managing Bridge Tables	343
	10.5 The Keyword Dimension	347
	Dealing with Hierarchies	351
	10.6 Maintaining Dimension Hierarchies	351
	10.7 Help for Hierarchies	355
	10.8 Five Alternatives for Better Employee Dimensional Modeling	359
	10.9 Alternate Hierarchies	365
	Customer Issues	366
	10.10 Dimension Embellishments	367
	10.11 Wrangling Behavior Tags	368
	10.12 Three Ways to Capture Customer Satisfaction	371
	Addresses and International Issues	374
	10.13 Think Globally, Act Locally	374
	10.14 Warehousing without Borders	378
	10.15 Spatially Enabling Your Data Warehouse	383
	10.16 Multinational Dimensional Data Warehouse Considerations	387
	Industry Scenarios and Idiosyncrasies	389
	10.17 An Insurance Data Warehouse Case Study	389
	10.18 Traveling through Databases	393
	10.19 Human Resources Dimensional Models	396
	10.20 Not So Fast	400
	10.21 The Budgeting Chain	403
	10.22 Compliance-Enabled Data Warehouses	407
	10.23 Clicking with Your Customer	409
	10.24 The Special Dimensions of the Clickstream	413
	10.25 Fact Tables for Text Document Searching	417
	10.26 Enabling Market Basket Analysis	420
11	Back Room ETL and Data Quality	425
	Planning the ETL System	425
	11.1 Surrounding the ETL Requirements	425

11.2 The 34 Subsystems of ETL	430
11.3 Doing the Work at Extract Time.	434
11.4 Is Data Staging Relational?	437
11.5 Staging Areas and ETL Tools	441
11.6 Should You Use an ETL Tool?	442
11.7 Document the ETL System	445
11.8 Measure Twice, Cut Once.	445
11.9 Brace for Incoming	449
11.10 Building a Change Data Capture System	452
Data Quality Considerations	454
11.11 Dealing with Dirty Data	454
11.12 An Architecture for Data Quality	460
11.13 Indicators of Quality	468
11.14 Is Your Data Correct?	471
11.15 Eight Recommendations for International Data Quality	474
11.16 Using Regular Expressions for Data Cleaning	477
Populating Fact and Dimension Tables	481
11.17 Pipelining Your Surrogates	481
11.18 Replicating Dimensions Correctly	485
11.19 Identify Dimension Changes Using Cyclic Redundancy Checksums	486
11.20 Maintaining Back Pointers to Operational Sources.	487
11.21 Creating Historical Dimension Rows	488
11.22 Backward in Time	491
11.23 Early-Arriving Facts	494
11.24 Slowly Changing Entities	495
11.25 Creating, Using, and Maintaining Junk Dimensions	497
11.26 Using the SQL MERGE for Slowly Changing Dimensions	499
11.27 Being Offline as Little as Possible	502
Supporting Real Time	503
11.28 Working in Web Time	503
11.29 Real-Time Partitions	507
11.30 The Real-Time Triage	510
12 Technical Architecture Considerations	513
Overall Technical/System Architecture	513
12.1 Can the Data Warehouse Benefit from SOA?	513

12.2 Picking the Right Approach to MDM	515
12.3 Building Custom Tools for the DW/BI System	520
12.4 Welcoming the Packaged App	522
12.5 ERP Vendors: Bring Down Those Walls	525
12.6 Building a Foundation for Smart Applications	528
12.7 RFID Tags and Smart Dust	533
Presentation Server Architecture	535
12.8 The Aggregate Navigator.	536
12.9 Aggregate Navigation with (Almost) No Metadata	539
12.10 Relating to OLAP	546
12.11 Dimensional Relational versus OLAP: The Final Deployment Conundrum	549
12.12 Dimensional Modeling for Microsoft Analysis Services	553
12.13 Architecting Your Data for Microsoft SQL Server 2005	554
12.14 Microsoft SQL Server Comes of Age for Data Warehousing.	556
Front Room Architecture.	560
12.15 The Second Revolution of User Interfaces	560
12.16 Designing the User Interface.	562
Metadata.	566
12.17 Meta Meta Data Data.	567
12.18 Creating the Metadata Strategy.	570
Infrastructure and Security Considerations.	572
12.19 Watching the Watchers.	572
12.20 Catastrophic Failure	576
12.21 Digital Preservation.	579
12.22 Creating the Advantages of a 64-Bit Server	582
12.23 Server Configuration Considerations	583
12.24 Adjust Your Thinking for SANs	585
13 Front Room Business Intelligence Applications.	589
Delivering Value with Business Intelligence	589
13.1 The Promise of Decision Support	589
13.2 Beyond Paving the Cow Paths.	593
13.3 Big Shifts Happening in BI	596
13.4 Behavior: The Next Marquee Application.	598

Implementing the Business Intelligence Layer	601
13.5 Think Like a Software Development Manager	601
13.6 Standard Reports: Basics for Business Users	602
13.7 Building and Delivering BI Reports.	607
13.8 The BI Portal	610
13.9 Dashboards Done Right	612
13.10 Don't Be Overly Reliant on Your Data Access Tool's Metadata	613
Mining Data to Uncover Relationships	615
13.11 Digging into Data Mining	615
13.12 Preparing for Data Mining	617
13.13 The Perfect Handoff	621
13.14 Get Started with Data Mining Now.	625
Dealing with SQL.	629
13.15 Simple Drill Across in SQL	629
13.16 The Problem with Comparisons	631
13.17 SQL Roadblocks and Pitfalls	635
13.18 Features for Query Tools	638
13.19 Turbocharge Your Query Tools	640
13.20 Smarter Data Warehouses.	644
14 Maintenance and Growth Considerations.	651
Deploying Successfully	651
14.1 Don't Forget the Owner's Manual.	651
14.2 Let's Improve Our Operating Procedures	655
14.3 Marketing the DW/BI System	656
14.4 Coping with Growing Pains	658
Sustaining for Ongoing Impact	661
14.5 Data Warehouse Checkups.	661
14.6 Boosting Business Acceptance.	667
14.7 Educate Management to Sustain DW/BI Success	670
14.8 Getting Your Data Warehouse Back on Track	673
14.9 Upgrading Your BI Architecture	674
14.10 Four Fixes for Legacy Data Warehouses.	676
14.11 A Data Warehousing Fitness Program for Lean Times.	680
Index of Articles	685
Index	693