

The Kimball Group Reader

Relentlessly Practical Tools for Data Warehousing and Business Intelligence

Remastered Collection

Ralph Kimball and Margy Ross
with Bob Becker, Joy Mundy, and Warren Thornthwaite

WILEY

Contents

Introduction	xxv
1 The Reader 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	9
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	25
1.9 Judge Your BI Tool through Your Dimensions	28
1.10 Fact Tables	31
1.11 Exploit Your Fact Tables	33
2 Before You Dive In	35
Before Data Warehousing	35
2.1 History Lesson on Ralph Kimball and Xerox PARC	36
Historical Perspective	37
2.2 The Database Market Splits	37
2.3 Bringing Up Supermarts	40
Dealing with Demanding Realities	47
2.4 Brave New Requirements for Data Warehousing	47
2.5 Coping with the Brave New Requirements	52
2.6 Stirring Things Up	57
2.7 Design Constraints and Unavoidable Realities	60

2.8 Two Powerful Ideas	64
2.9 Data Warehouse Dining Experience	67
2.10 Easier Approaches for Harder Problems	70
2.11 Expanding Boundaries of the Data Warehouse	72
3 Project/Program Planning	75
Professional Responsibilities	75
3.1 Professional Boundaries	75
3.2 An Engineer’s View	78
3.3 Beware the Objection Removers	82
3.4 What Does the Central Team Do?	86
3.5 Avoid Isolating DW and BI Teams.	90
3.6 Better Business Skills for BI and Data Warehouse Professionals	91
3.7 Risky Project Resources Are Risky Business	93
3.8 Implementation Analysis Paralysis	95
3.9 Contain DW/BI Scope Creep and Avoid Scope Theft	96
3.10 Are IT Procedures Beneficial to DW/BI Projects?	98
Justification and Sponsorship	100
3.11 Habits of Effective Sponsors	100
3.12 TCO Starts with the End User	103
Kimball Methodology.	108
3.13 Kimball Lifecycle in a Nutshell	108
3.14 Off the Bench.	111
3.15 The Anti-Architect	112
3.16 Think Critically When Applying Best Practices	115
3.17 Eight Guidelines for Low Risk Enterprise Data Warehousing	118
4 Requirements Definition	123
Gathering Requirements.	123
4.1 Alan Alda’s Interviewing Tips for Uncovering Business Requirements	123
4.2 More Business Requirements Gathering Dos and Don’ts	127
4.3 Balancing Requirements and Realities	129
4.4 Overcoming Obstacles When Gathering Business Requirements	130
4.5 Surprising Value of Data Profiling.	133

Organizing around Business Processes	134
4.6 Focus on Business Processes, Not Business Departments!	134
4.7 Identifying Business Processes	135
4.8 Business Process Decoder Ring.	137
4.9 Relationship between Strategic Business Initiatives and Business Processes.	138
Wrapping Up the Requirements	139
4.10 The Bottom-Up Misnomer	140
4.11 Think Dimensionally (Beyond Data Modeling)	144
4.12 Using the Dimensional Model to Validate Business Requirements.	145
5 Data Architecture	147
Making the Case for Dimensional Modeling.	147
5.1 Is ER Modeling Hazardous to DSS?	147
5.2 A Dimensional Modeling Manifesto	151
5.3 There Are No Guarantees.	159
Enterprise Data Warehouse Bus Architecture.	163
5.4 Divide and Conquer.	163
5.5 The Matrix.	166
5.6 The Matrix: Revisited	170
5.7 Drill Down into a Detailed Bus Matrix	174
Agile Project Considerations	176
5.8 Relating to Agile Methodologies	176
5.9 Is Agile Enterprise Data Warehousing an Oxymoron?	177
5.10 Going Agile? Start with the Bus Matrix.	179
5.11 Conformed Dimensions as the Foundation for Agile Data Warehousing	180
Integration Instead of Centralization.	181
5.12 Integration for Real People	181
5.13 Build a Ready-to-Go Resource for Enterprise Dimensions	185
5.14 Data Stewardship 101: The First Step to Quality and Consistency	186
5.15 To Be or Not To Be Centralized.	189
Contrast with the Corporate Information Factory	192
5.16 Differences of Opinion	193
5.17 Much Ado about Nothing	198

5.18 Don't Support Business Intelligence with a Normalized EDW	199
5.19 Complementing 3NF EDWs with Dimensional Presentation Areas	201
6 Dimensional Modeling Fundamentals	203
Basics of Dimensional Modeling	203
6.1 Fact Tables and Dimension Tables	203
6.2 Drilling Down, Up, and Across	207
6.3 The Soul of the Data Warehouse, Part One: Drilling Down	210
6.4 The Soul of the Data Warehouse, Part Two: Drilling Across	213
6.5 The Soul of the Data Warehouse, Part Three: Handling Time	216
6.6 Graceful Modifications to Existing Fact and Dimension Tables	219
Dos and Don'ts	220
6.7 Kimball's Ten Essential Rules of Dimensional Modeling	221
6.8 What Not to Do	223
Myths about Dimensional Modeling	226
6.9 Dangerous Preconceptions	226
6.10 Fables and Facts	228
7 Dimensional Modeling Tasks and Responsibilities	233
Design Activities	233
7.1 Letting the Users Sleep	233
7.2 Practical Steps for Designing a Dimensional Model	240
7.3 Staffing the Dimensional Modeling Team	243
7.4 Involve Business Representatives in Dimensional Modeling	244
7.5 Managing Large Dimensional Design Teams	246
7.6 Use a Design Charter to Keep Dimensional Modeling Activities on Track	248
7.7 The Naming Game	249
7.8 What's in a Name?	250
7.9 When Is the Dimensional Design Done?	253
Design Review Activities	254
7.10 Design Review Dos and Don'ts	255
7.11 Fistful of Flaws	257
7.12 Rating Your Dimensional Data Warehouse	260

8	Fact Table Core Concepts	267
	Granularity	267
	8.1 Declaring the Grain	267
	8.2 Keep to the Grain in Dimensional Modeling	270
	8.3 Warning: Summary Data May Be Hazardous to Your Health	272
	8.4 No Detail Too Small	273
	Types of Fact Tables	276
	8.5 Fundamental Grains	277
	8.6 Modeling a Pipeline with an Accumulating Snapshot	280
	8.7 Combining Periodic and Accumulating Snapshots	282
	8.8 Complementary Fact Table Types	284
	8.9 Modeling Time Spans	286
	8.10 A Rolling Prediction of the Future, Now and in the Past	289
	8.11 Timespan Accumulating Snapshot Fact Tables	293
	8.12 Is it a Dimension, a Fact, or Both?	294
	8.13 Factless Fact Tables	295
	8.14 Factless Fact Tables? Sounds Like Jumbo Shrimp?	298
	8.15 What Didn't Happen	299
	8.16 Factless Fact Tables for Simplification	302
	Parent-Child Fact Tables	304
	8.17 Managing Your Parents	304
	8.18 Patterns to Avoid When Modeling Header/Line Item Transactions	307
	Fact Table Keys and Degenerate Dimensions	309
	8.19 Fact Table Surrogate Keys	309
	8.20 Reader Suggestions on Fact Table Surrogate Keys	310
	8.21 Another Look at Degenerate Dimensions	312
	8.22 Creating a Reference Dimension for Infrequently Accessed Degenerates	313
	Miscellaneous Fact Table Design Patterns	314
	8.23 Put Your Fact Tables on a Diet	314
	8.24 Keeping Text Out of the Fact Table	316
	8.25 Dealing with Nulls in a Dimensional Model	317
	8.26 Modeling Data as Both a Fact and Dimension Attribute	318

8.27	When a Fact Table Can Be Used as a Dimension Table	319
8.28	Sparse Facts and Facts with Short Lifetimes	321
8.29	Pivoting the Fact Table with a Fact Dimension	323
8.30	Accumulating Snapshots for Complex Workflows	324
9	Dimension Table Core Concepts	327
	Dimension Table Keys.	327
9.1	Surrogate Keys	327
9.2	Keep Your Keys Simple	331
9.3	Durable “Super-Natural” Keys	333
	Date and Time Dimension Considerations	334
9.4	It’s Time for Time.	335
9.5	Surrogate Keys for the Time Dimension	337
9.6	Latest Thinking on Time Dimension Tables.	339
9.7	Smart Date Keys to Partition Fact Tables.	341
9.8	Updating the Date Dimension	342
9.9	Handling All the Dates	343
	Miscellaneous Dimension Patterns	345
9.10	Selecting Default Values for Nulls	345
9.11	Data Warehouse Role Models	347
9.12	Mystery Dimensions	350
9.13	De-Clutter with Junk Dimensions	353
9.14	Showing the Correlation between Dimensions	354
9.15	Causal (Not Casual) Dimensions	356
9.16	Resist Abstract Generic Dimensions	359
9.17	Hot-Swappable Dimensions	360
9.18	Accurate Counting with a Dimensional Supplement	361
	Slowly Changing Dimensions	363
9.19	Perfectly Partitioning History with Type 2 SCD	363
9.20	Many Alternate Realities.	364
9.21	Monster Dimensions	367
9.22	When a Slowly Changing Dimension Speeds Up	370
9.23	When Do Dimensions Become Dangerous?	372
9.24	Slowly Changing Dimensions Are Not Always as Easy as 1, 2, and 3.	373

9.25 Slowly Changing Dimension Types 0, 4, 5, 6 and 7	378
9.26 Dimension Row Change Reason Attributes.	382
10 More Dimension Patterns and Considerations	385
Snowflakes, Outriggers, and Bridges	385
10.1 Snowflakes, Outriggers, and Bridges	385
10.2 A Trio of Interesting Snowflakes	388
10.3 Help for Dimensional Modeling	392
10.4 Managing Bridge Tables.	395
10.5 The Keyword Dimension	399
10.6 Potential Bridge (Table) Detours.	403
10.7 Alternatives for Multi-Valued Dimensions	405
10.8 Adding a Mini-Dimension to a Bridge Table.	407
Dealing with Hierarchies.	409
10.9 Maintaining Dimension Hierarchies	409
10.10 Help for Hierarchies	414
10.11 Five Alternatives for Better Employee Dimensional Modeling	417
10.12 Avoiding Alternate Organization Hierarchies.	425
10.13 Alternate Hierarchies.	426
Customer Issues	427
10.14 Dimension Embellishments.	427
10.15 Wrangling Behavior Tags	429
10.16 Three Ways to Capture Customer Satisfaction	431
10.17 Extreme Status Tracking for Real-Time Customer Analysis.	435
Addresses and International Issues	439
10.18 Think Globally, Act Locally	439
10.19 Warehousing without Borders.	443
10.20 Spatially Enabling Your Data Warehouse.	448
10.21 Multinational Dimensional Data Warehouse Considerations	452
Industry Scenarios and Idiosyncrasies	453
10.22 Industry Standard Data Models Fall Short	453
10.23 An Insurance Data Warehouse Case Study	455
10.24 Traveling through Databases	460
10.25 Human Resources Dimensional Models	463

10.26	Managing Backlogs Dimensionally	467
10.27	Not So Fast	468
10.28	The Budgeting Chain	471
10.29	Compliance-Enabled Data Warehouses	475
10.30	Clicking with Your Customer	477
10.31	The Special Dimensions of the Clickstream.	482
10.32	Fact Tables for Text Document Searching	485
10.33	Enabling Market Basket Analysis	489
11	Back Room ETL and Data Quality	495
	Planning the ETL System	495
	11.1 Surrounding the ETL Requirements.	495
	11.2 The 34 Subsystems of ETL	500
	11.3 Six Key Decisions for ETL Architectures	504
	11.4 Three ETL Compromises to Avoid	508
	11.5 Doing the Work at Extract Time	510
	11.6 Is Data Staging Relational?	513
	11.7 Staging Areas and ETL Tools	517
	11.8 Should You Use an ETL Tool?	518
	11.9 Call to Action for ETL Tool Providers	521
	11.10 Document the ETL System	522
	11.11 Measure Twice, Cut Once	523
	11.12 Brace for Incoming	527
	11.13 Building a Change Data Capture System.	530
	11.14 Disruptive ETL Changes	531
	11.15 New Directions for ETL	533
	Data Quality Considerations.	535
	11.16 Dealing With Data Quality: Don't Just Sit There, Do Something!.	535
	11.17 Data Warehouse Testing Recommendations	537
	11.18 Dealing with Dirty Data	539
	11.19 An Architecture for Data Quality	545
	11.20 Indicators of Quality: The Audit Dimension	553
	11.21 Adding an Audit Dimension to Track Lineage and Confidence	556
	11.22 Add Uncertainty to Your Fact Table	559

11.23 Have You Built Your Audit Dimension Yet?	560
11.24 Is Your Data Correct?	562
11.25 Eight Recommendations for International Data Quality	565
11.26 Using Regular Expressions for Data Cleaning	568
Populating Fact and Dimension Tables	572
11.27 Pipelining Your Surrogates	572
11.28 Unclogging the Fact Table Surrogate Key Pipeline	576
11.29 Replicating Dimensions Correctly	579
11.30 Identify Dimension Changes Using Cyclic Redundancy Checksums	580
11.31 Maintaining Back Pointers to Operational Sources	581
11.32 Creating Historical Dimension Rows	582
11.33 Facing the Re-Keying Crisis	585
11.34 Backward in Time	587
11.35 Early-Arriving Facts	590
11.36 Slowly Changing Entities	591
11.37 Using the SQL MERGE Statement for Slowly Changing Dimensions	593
11.38 Creating and Managing Shrunk Dimensions	595
11.39 Creating and Managing Mini-Dimensions	597
11.40 Creating, Using, and Maintaining Junk Dimensions	599
11.41 Building Bridges	601
11.42 Being Offline as Little as Possible	605
Supporting Real Time	606
11.43 Working in Web Time	606
11.44 Real-Time Partitions	610
11.45 The Real-Time Triage	613
12 Technical Architecture Considerations	617
Overall Technical/System Architecture	617
12.1 Can the Data Warehouse Benefit from SOA?	617
12.2 Picking the Right Approach to MDM	619
12.3 Building Custom Tools for the DW/BI System	625
12.4 Welcoming the Packaged App	626
12.5 ERP Vendors: Bring Down Those Walls	629
12.6 Building a Foundation for Smart Applications	632

12.7 RFID Tags and Smart Dust	637
12.8 Is Big Data Compatible with the Data Warehouse?	640
12.9 The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics	641
12.10 Newly Emerging Best Practices for Big Data	659
12.11 The Hyper-Granular Active Archive	670
Presentation Server Architecture.	672
12.12 Columnar Databases: Game Changers for DW/BI Deployment	672
12.13 There Is no Database Magic	673
12.14 Relating to OLAP	676
12.15 Dimensional Relational versus OLAP: The Final Deployment Conundrum	679
12.16 Microsoft SQL Server Comes of Age for Data Warehousing.	682
12.17 The Aggregate Navigator	686
12.18 Aggregate Navigation with (Almost) No Metadata	690
Front Room Architecture	697
12.19 The Second Revolution of User Interfaces	697
12.20 Designing the User Interface	700
Metadata	704
12.21 Meta Meta Data Data.	704
12.22 Creating the Metadata Strategy	708
12.23 Leverage Process Metadata for Self-Monitoring DW Operations.	709
Infrastructure and Security Considerations	712
12.24 Watching the Watchers	712
12.25 Catastrophic Failure	716
12.26 Digital Preservation	719
12.27 Creating the Advantages of a 64-Bit Server	722
12.28 Server Configuration Considerations	723
12.29 Adjust Your Thinking for SANs	726
13 Front Room Business Intelligence Applications	729
Delivering Value with Business Intelligence	729
13.1 The Promise of Decision Support	730
13.2 Beyond Paving the Cow Paths	733
13.3 BI Components for Business Value	736

13.4 Big Shifts Happening in BI	738
13.5 Behavior: The Next Marquee Application.	740
Implementing the Business Intelligence Layer.	743
13.6 Three Critical Components for Successful Self-Service BI	743
13.7 Leverage Data Visualization Tools, But Avoid Anarchy.	745
13.8 Think Like a Software Development Manager	747
13.9 Standard Reports: Basics for Business Users	748
13.10 Building and Delivering BI Reports	753
13.11 The BI Portal.	757
13.12 Dashboards Done Right	759
13.13 Don't Be Overly Reliant on Your Data Access Tool's Metadata	760
13.14 Making Sense of the Semantic Layer.	762
Mining Data to Uncover Relationships	764
13.15 Digging into Data Mining.	764
13.16 Preparing for Data Mining	766
13.17 The Perfect Handoff	770
13.18 Get Started with Data Mining Now	774
13.19 Leverage Your Dimensional Model for Predictive Analytics	778
13.20 Does Your Organization Need an Analytic Sandbox?	779
Dealing with SQL	781
13.21 Simple Drill Across in SQL.	781
13.22 An Excel Macro for Drilling Across	783
13.23 The Problem with Comparisons	785
13.24 SQL Roadblocks and Pitfalls	789
13.25 Features for Query Tools.	792
13.26 Turbocharge Your Query Tools.	794
13.27 Smarter Data Warehouses	798
14 Maintenance and Growth Considerations	805
Deploying Successfully	805
14.1 Don't Forget the Owner's Manual	805
14.2 Let's Improve Our Operating Procedures	809
14.3 Marketing the DW/BI System	811
14.4 Coping with Growing Pains	812

- Sustaining for Ongoing Impact 816
 - 14.5 Data Warehouse Checkups. 816
 - 14.6 Boosting Business Acceptance 822
 - 14.7 Educate Management to Sustain DW/BI Success 825
 - 14.8 Getting Your Data Warehouse Back on Track 828
 - 14.9 Upgrading Your BI Architecture 829
 - 14.10 Four Fixes for Legacy Data Warehouses 831
 - 14.11 A Data Warehousing Fitness Program for Lean Times 835
 - 14.12 Enjoy the Sunset. 839
- 15 Final Thoughts 841**
 - Key Insights and Reminders 841
 - 15.1 Final Word of the Day: Collaboration 841
 - 15.2 Tried and True Concepts for DW/BI Success 843
 - 15.3 Key Tenets of the Kimball Method 845
 - A Look to the Future. 847
 - 15.4 The Future Is Bright 847
- Article Index 853
- Index 861