

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition

[Ralph Kimball](#), [Margy Ross](#)

ISBN: 978-1-118-53080-1

TABLE OF CONTENTS

Introduction xxvii

1 Data Warehousing, Business Intelligence, and Dimensional Modeling Primer 1

Different Worlds of Data Capture and Data Analysis 2
Goals of Data Warehousing and Business Intelligence 3
Dimensional Modeling Introduction 7
Kimball's DW/BI Architecture 18
Alternative DW/BI Architectures 26
Dimensional Modeling Myths 30
More Reasons to Think Dimensionally 32
Agile Considerations 34
Summary 35

2 Kimball Dimensional Modeling Techniques Overview 37

Fundamental Concepts 37
Basic Fact Table Techniques 41
Basic Dimension Table Techniques 46
Integration via Conformed Dimensions 50
Dealing with Slowly Changing Dimension Attributes 53
Dealing with Dimension Hierarchies 56
Advanced Fact Table Techniques 58
Advanced Dimension Techniques 62
Special Purpose Schemas 67

3 Retail Sales 69

Four-Step Dimensional Design Process 70
Retail Case Study 72
Dimension Table Details 79
Retail Schema in Action 94
Retail Schema Extensibility 95
Factless Fact Tables 97

Dimension and Fact Table Keys 98
Resisting Normalization Urges 104
Summary 109

4 Inventory 111

Value Chain Introduction 111
Inventory Models 112
Fact Table Types 119
Value Chain Integration 122
Enterprise Data Warehouse Bus Architecture 123
Conformed Dimensions 130
Conformed Facts 138
Summary 139

5 Procurement 141

Procurement Case Study 141
Procurement Transactions and Bus Matrix 142
Slowly Changing Dimension Basics 147
Hybrid Slowly Changing Dimension Techniques 159
Slowly Changing Dimension Recap 164
Summary 165

6 Order Management 167

Order Management Bus Matrix 168
Order Transactions 168
Invoice Transactions 187
Accumulating Snapshot for Order Fulfillment Pipeline 194
Summary 199

7 Accounting 201

Accounting Case Study and Bus Matrix 202
General Ledger Data 203
Budgeting Process 210
Dimension Attribute Hierarchies 214
Consolidated Fact Tables 224
Role of OLAP and Packaged Analytic Solutions 226
Summary 227

8 Customer Relationship Management 229

- CRM Overview 230
- Customer Dimension Attributes 233
- Bridge Tables for Multivalued Dimensions 245
- Complex Customer Behavior 249
- Customer Data Integration Approaches 256
- Low Latency Reality Check 260
- Summary 261

9 Human Resources Management 263

- Employee Profile Tracking 263
- Headcount Periodic Snapshot 267
- Bus Matrix for HR Processes 268
- Packaged Analytic Solutions and Data Models 270
- Recursive Employee Hierarchies 271
- Multivalued Skill Keyword Attributes 274
- Survey Questionnaire Data 277
- Summary 279

10 Financial Services 281

- Banking Case Study and Bus Matrix 282
- Dimension Triage to Avoid Too Few Dimensions 283
- Supertype and Subtype Schemas for Heterogeneous Products 293
- Hot Swappable Dimensions 296
- Summary 296

11 Telecommunications 297

- Telecommunications Case Study and Bus Matrix 297
- General Design Review Considerations 299
- Design Review Guidelines 304
- Draft Design Exercise Discussion 306
- Remodeling Existing Data Structures 309
- Geographic Location Dimension 310
- Summary 310

12 Transportation 311

- Airline Case Study and Bus Matrix 311
- Extensions to Other Industries 317
- Combining Correlated Dimensions 318

More Date and Time Considerations 321

Localization Recap 324

Summary 324

13 Education 325

University Case Study and Bus Matrix 325

Accumulating Snapshot Fact Tables 326

Factless Fact Tables 329

More Educational Analytic Opportunities 336

Summary 336

14 Healthcare 339

Healthcare Case Study and Bus Matrix 339

Claims Billing and Payments 342

Electronic Medical Records 348

Facility/Equipment Inventory Utilization 351

Dealing with Retroactive Changes 351

Summary 352

15 Electronic Commerce 353

Clickstream Source Data 353

Clickstream Dimensional Models 357

Integrating Clickstream into Web Retailer's Bus Matrix 368

Profitability Across Channels Including Web 370

Summary 373

16 Insurance 375

Insurance Case Study 376

Policy Transactions 379

Premium Periodic Snapshot 385

More Insurance Case Study Background 388

Claim Transactions 390

Claim Accumulating Snapshot 392

Policy/Claim Consolidated Periodic Snapshot 395

Factless Accident Events 396

Common Dimensional Modeling Mistakes to Avoid 397

Summary 401

17 Kimball DW/BI Lifecycle Overview 403

Lifecycle Roadmap 404
Lifecycle Launch Activities 406
Lifecycle Technology Track 416
Lifecycle Data Track 420
Lifecycle BI Applications Track 422
Lifecycle Wrap-up Activities 424
Common Pitfalls to Avoid 426
Summary 427

18 Dimensional Modeling Process and Tasks 429

Modeling Process Overview 429
Get Organized 431
Design the Dimensional Model 434
Summary 441

19 ETL Subsystems and Techniques 443

Round Up the Requirements 444
The 34 Subsystems of ETL 449
Extracting: Getting Data into the Data Warehouse 450
Cleaning and Conforming Data 455
Delivering: Prepare for Presentation 463
Managing the ETL Environment 483
Summary 496

20 ETL System Design and Development Process and Tasks 497

ETL Process Overview 497
Develop the ETL Plan 498
Develop One-Time Historic Load Processing 503
Develop Incremental ETL Processing 512
Real-Time Implications 520
Summary 526

21 Big Data Analytics 527

Big Data Overview 527
Recommended Best Practices for Big Data 531
Summary 542
Index 543