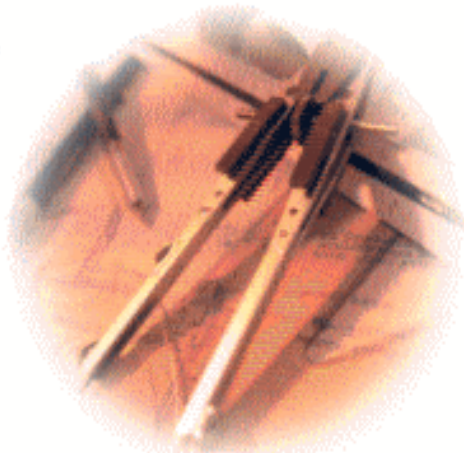


The Data Warehouse Toolkit

Second Edition

The Complete
Guide to
Dimensional
Modeling

Ralph Kimball
Margy Ross



Acknowledgments	xv
Introduction	xvii
 Chapter 1 Dimensional Modeling Primer	 1
Different Information Worlds	2
Goals of a Data Warehouse	2
The Publishing Metaphor	4
Components of a Data Warehouse	6
Operational Source Systems	7
Data Staging Area	8
Data Presentation	10
Data Access Tools	13
Additional Considerations	14
Dimensional Modeling Vocabulary	16
Fact Table	16
Dimension Tables	19
Bringing Together Facts and Dimensions	21
Dimensional Modeling Myths	24
Common Pitfalls to Avoid	26
Summary	27
 Chapter 2 Retail Sales	 29
Four-Step Dimensional Design Process	30
Retail Case Study	32
Step 1. Select the Business Process	33
Step 2. Declare the Grain	34
Step 3. Choose the Dimensions	35
Step 4. Identify the Facts	36

Dimension Table Attributes	38
Date Dimension	38
Product Dimension	42
Store Dimension	45
Promotion Dimension	46
Degenerate Transaction Number Dimension	50
Retail Schema in Action	51
Retail Schema Extensibility	52
Resisting Comfort Zone Urges	54
Dimension Normalization (Snowflaking)	55
Too Many Dimensions	57
Surrogate Keys	58
Market Basket Analysis	62
Summary	65
Chapter 3 Inventory	67
Introduction to the Value Chain	68
Inventory Models	69
Inventory Periodic Snapshot	69
Inventory Transactions	74
Inventory Accumulating Snapshot	75
Value Chain Integration	76
Data Warehouse Bus Architecture	78
Data Warehouse Bus Matrix	79
Conformed Dimensions	82
Conformed Facts	87
Summary	88
Chapter 4 Procurement	89
Procurement Case Study	89
Procurement Transactions	90
Multiple- versus Single-Transaction Fact Tables	91
Complementary Procurement Snapshot	93

Slowly Changing Dimensions	95
Type 1: Overwrite the Value	95
Type 2: Add a Dimension Row	97
Type 3: Add a Dimension Column	100
Hybrid Slowly Changing Dimension Techniques	102
Predictable Changes with Multiple Version Overlays	102
Unpredictable Changes with Single Version Overlay	103
More Rapidly Changing Dimensions	105
Summary	105

Chapter 5	Order Management	107
	Introduction to Order Management	108
	Order Transactions	109
	Fact Normalization	109
	Dimension Role-Playing	110
	Product Dimension Revisited	111
	Customer Ship-To Dimension	113
	Deal Dimension	116
	Degenerate Dimension for Order Number	117
	Junk Dimensions	117
	Multiple Currencies	119
	Header and Line Item Facts with Different Granularity	121
	Invoice Transactions	122
	Profit and Loss Facts	124
	Profitability—The Most Powerful Data Mart	126
	Profitability Words of Warning	127
	Customer Satisfaction Facts	127
	Accumulating Snapshot for the Order Fulfillment Pipeline	128
	Lag Calculations	130
	Multiple Units of Measure	130
	Beyond the Rear-View Mirror	132
	Fact Table Comparison	132
	Transaction Fact Tables	133
	Periodic Snapshot Fact Tables	134
	Accumulating Snapshot Fact Tables	134

Designing Real-Time Partitions	135
Requirements for the Real-Time Partition	136
Transaction Grain Real-Time Partition	136
Periodic Snapshot Real-Time Partition	137
Accumulating Snapshot Real-Time Partition	138
Summary	139
Chapter 6 Customer Relationship Management	141
CRM Overview	142
Operational and Analytical CRM	143
Packaged CRM	145
Customer Dimension	146
Name and Address Parsing	147
Other Common Customer Attributes	150
Dimension Outriggers for a Low-Cardinality Attribute Set	153
Large Changing Customer Dimensions	154
Implications of Type 2 Customer Dimension Changes	159
Customer Behavior Study Groups	160
Commercial Customer Hierarchies	161
Combining Multiple Sources of Customer Data	168
Analyzing Customer Data from Multiple Business Processes	169
Summary	170
Chapter 7 Accounting	173
Accounting Case Study	174
General Ledger Data	175
General Ledger Periodic Snapshot	175
General Ledger Journal Transactions	177
Financial Statements	180
Budgeting Process	180
Consolidated Fact Tables	184
Role of OLAP and Packaged Analytic Solutions	185
Summary	186

Chapter 8	Human Resources Management	187
	Time-Stamped Transaction Tracking in a Dimension	188
	Time-Stamped Dimension with Periodic Snapshot Facts	191
	Audit Dimension	193
	Keyword Outrigger Dimension	194
	AND/OR Dilemma	195
	Searching for Substrings	196
	Survey Questionnaire Data	197
	Summary	198
Chapter 9	Financial Services	199
	Banking Case Study	200
	Dimension Triage	200
	Household Dimension	204
	Multivalued Dimensions	205
	Minidimensions Revisited	206
	Arbitrary Value Banding of Facts	207
	Point-in-Time Balances	208
	Heterogeneous Product Schemas	210
	Heterogeneous Products with Transaction Facts	215
	Summary	215
Chapter 10	Telecommunications and Utilities	217
	Telecommunications Case Study	218
	General Design Review Considerations	220
	Granularity	220
	Date Dimension	222
	Degenerate Dimensions	222
	Dimension Decodes and Descriptions	222
	Surrogate Keys	223
	Too Many (or Too Few) Dimensions	223
	Draft Design Exercise Discussion	223
	Geographic Location Dimension	226
	Location Outrigger	226
	Leveraging Geographic Information Systems	227
	Summary	227

Chapter 11	Transportation	229
	Airline Frequent Flyer Case Study	230
	Multiple Fact Table Granularities	230
	Linking Segments into Trips	233
	Extensions to Other Industries	234
	Cargo Shipper	234
	Travel Services	235
	Combining Small Dimensions into a Superdimension	236
	Class of Service	236
	Origin and Destination	237
	More Date and Time Considerations	239
	Country-Specific Calendars	239
	Time of Day as a Dimension or Fact	240
	Date and Time in Multiple Time Zones	240
	Summary	241
Chapter 12	Education	243
	University Case Study	244
	Accumulating Snapshot for Admissions Tracking	244
	Factless Fact Tables	246
	Student Registration Events	247
	Facilities Utilization Coverage	249
	Student Attendance Events	250
	Other Areas of Analytic Interest	253
	Summary	254
Chapter 13	Health Care	255
	Health Care Value Circle	256
	Health Care Bill	258
	Roles Played By the Date Dimension	261
	Multivalued Diagnosis Dimension	262
	Extending a Billing Fact Table to Show Profitability	265
	Dimensions for Billed Hospital Stays	266

Complex Health Care Events	267
Medical Records	269
Fact Dimension for Sparse Facts	269
Going Back in Time	271
Late-Arriving Fact Rows	271
Late-Arriving Dimension Rows	273
Summary	274

Chapter 14 Electronic Commerce	277
Web Client-Server Interactions Tutorial	278
Why the Clickstream Is Not Just Another Data Source	281
Challenges of Tracking with Clickstream Data	282
Specific Dimensions for the Clickstream	287
Clickstream Fact Table for Complete Sessions	292
Clickstream Fact Table for Individual Page Events	295
Aggregate Clickstream Fact Tables	298
Integrating the Clickstream Data Mart into the Enterprise Data Warehouse	299
Electronic Commerce Profitability Data Mart	300
Summary	303

Chapter 15 Insurance	305
Insurance Case Study	306
Insurance Value Chain	307
Draft Insurance Bus Matrix	309
Policy Transactions	309
Dimension Details and Techniques	310
Alternative (or Complementary) Policy Accumulating Snapshot	315
Policy Periodic Snapshot	316
Conformed Dimensions	316
Conformed Facts	316
Heterogeneous Products Again	318
Multivalued Dimensions Again	318

More Insurance Case Study Background	319
Updated Insurance Bus Matrix	320
Claims Transactions	322
Claims Accumulating Snapshot	323
Policy/Claims Consolidated Snapshot	324
Factless Accident Events	325
Common Dimensional Modeling Mistakes to Avoid	326
Summary	330
Chapter 16 Building the Data Warehouse	331
Business Dimensional Lifecycle Road Map	332
Road Map Major Points of Interest	333
Project Planning and Management	334
Assessing Readiness	334
Scoping	336
Justification	336
Staffing	337
Developing and Maintaining the Project Plan	339
Business Requirements Definition	340
Requirements Preplanning	341
Collecting the Business Requirements	343
Postcollection Documentation and Follow-up	345
Lifecycle Technology Track	347
Technical Architecture Design	348
Eight-Step Process for Creating the Technical Architecture	348
Product Selection and Installation	351
Lifecycle Data Track	353
Dimensional Modeling	353
Physical Design	355
Aggregation Strategy	356
Initial Indexing Strategy	357
Data Staging Design and Development	358
Dimension Table Staging	358
Fact Table Staging	361

Lifecycle Analytic Applications Track	362
Analytic Application Specification	363
Analytic Application Development	363
Deployment	364
Maintenance and Growth	365
Common Data Warehousing Mistakes to Avoid	366
Summary	369
Chapter 17 Present Imperatives and Future Outlook	371
Ongoing Technology Advances	372
Political Forces Demanding Security and Affecting Privacy	375
Conflict between Beneficial Uses and Insidious Abuses	375
Who Owns Your Personal Data?	376
What Is Likely to Happen? Watching the Watchers . . .	377
How Watching the Watchers Affects Data Warehouse Architecture	378
Designing to Avoid Catastrophic Failure	379
Catastrophic Failures	380
Countering Catastrophic Failures	380
Intellectual Property and Fair Use	383
Cultural Trends in Data Warehousing	383
Managing by the Numbers across the Enterprise	383
Increased Reliance on Sophisticated Key Performance Indicators	384
Behavior Is the New Marquee Application	385
Packaged Applications Have Hit Their High Point	385
Application Integration Has to Be Done by Someone	386
Data Warehouse Outsourcing Needs a Sober Risk Assessment	386
In Closing	387
Glossary	389
Index	419