Contents

Foreword xix
Acknowledgments xxi
Introduction xxiii

Chapter 1  **Fundamentals of Aggregates**  
Star Schema Basics 1  
  Operational Systems and the Data Warehouse 2  
    Operational Systems 3  
    Data Warehouse Systems 4  
Facts and Dimensions 5  
The Star Schema 7  
  Dimension Tables and Surrogate Keys 7  
  Fact Tables and Grain 10  
Using the Star Schema 13  
Multiple Stars and Conformance 15  
Data Warehouse Architecture 20  
Invisible Aggregates 22  
Improving Performance 23  
The Base Schema and the Aggregate Schema 25  
The Aggregate Navigator 26  
Principles of Aggregation 27  
  Providing the Same Results 27  
  The Same Facts and Dimension Attributes as the Base Schema 28
Design Principles for the Aggregate Schema 81
  A Separate Star for Each Aggregation 81
  Single Schema and the Level Field 81
  Drawbacks to the Single Schema Approach 84
  Advantages of Separate Tables 85
  Pre-Joined Aggregates 86
Naming Conventions 87
  Naming the Attributes 87
  Naming Aggregate Tables 88
Aggregate Dimension Design 90
  Attributes of Aggregate Dimensions 90
  Sourcing Aggregate Dimensions 91
  Shared Dimensions 92
Aggregate Fact Table Design 93
  Aggregate Facts: Names and Data Types 94
  No New Facts, Including Counts 94
  Degenerate Dimensions 96
  Audit Dimension 96
  Sourcing Aggregate Fact Tables 97
Pre-Joined Aggregate Design 98
Documenting the Aggregate Schema 98
  Identify Schema Families 99
  Identify Dimensional Conformance 99
  Documenting Aggregate Dimension Tables 101
  Documenting Aggregate Fact Tables 103
  Pre-Joined Aggregates 106
  Materialized Views and Materialized Query Tables 108
Summary 108

Chapter 4  Using Aggregates 109
Which Tables to Use? 110
  The Schema Design 110
  Relative Size 113
  Aggregate Portfolio and Availability 114
Requirements for the Aggregate Navigator 116
  Why an Aggregate Navigator? 116
  Two Views and Query Rewrite 117
  Dynamic Availability 120
  Multiple Front Ends 121
  Multiple Back Ends 123
Evaluating Aggregate Navigators 126
  Front-End Aggregate Navigators 127
    Approach 127
    Pros and Cons 128
Chapter 6  ETL Part 2: Loading Aggregates  187

The Source Data for Aggregate Tables  188
   Changed Data Identification  188
   Elimination of Redundant Processing  189
   Ensuring Conformance  190
   Loading the Base Schema and Aggregates Simultaneously  192

Loading Aggregate Dimensions  193
   Requirements for the Aggregate Dimension Load Process  194
   Extracting and Preparing the Records  195
   Identifying and Processing New Records  197
   Identifying and Processing Type 1 Changes  198
   Processing Type 2 Changes  203
   Key Mapping  204

Loading Aggregate Fact Tables  205
   Requirements for Loading Aggregate Fact Tables  205
   Acquire Data and Assemble Facts  205
      Selecting Source Columns  206
      Processing New Facts Only  208
      Calculating and Aggregating Facts  208
      One Query Does It All  209
   Identification of Surrogate Keys  210
   Aggregating Over Time  212

Dropping and Rebuilding Aggregates  214
   Dropping and Rebuilding Aggregate Dimension Tables  214
   Dropping and Rebuilding Aggregate Fact Tables  216

Pre-Joined Aggregates  217
   Dropping and Rebuilding a Pre-Joined Aggregate  217
   Incrementally Loading a Pre-Joined Aggregate  219

Materialized Views and Materialized Query Tables  221
   Defining Attributes for Aggregate Dimensions  221
   Optimizing the Hierarchy  222

Summary  223
## Chapter 7 Aggregates and Your Project 225

Data Warehouse Implementation 226
- Incremental Implementation of the Data Warehouse 226
- Planning Data Marts Around Conformed Dimensions 226
- Other Approaches 230
- Incorporating Aggregates into the Project 230
  - Aggregates and the First Data Mart 231
  - Subsequent Subject Areas 232

The Aggregate Project 233
- Strategy Stage 234
  - Technology Selection: Choosing an Aggregate Navigator 234
  - Additional Strategic Tasks and Deliverables 240
- Design Stage 241
  - Design of the Aggregate Schema and Load Specification 243
  - Design Documentation 243
  - Developing Test Plans for Aggregates 244
- Build Stage 245
  - Iterative Build and Aggregates 245
  - Build Tasks and Aggregates 246
- Deployment 250
  - Transitioning to Production, Final Testing, and Documentation 250
  - End User Education 252

Management of Aggregates 252
- Maintenance Responsibilities 252
- Ad Hoc Changes to Aggregate Portfolio 254
- An Ongoing Process 254

Summary 255

## Chapter 8 Advanced Aggregate Design 257

Aggregating Facts 258
- Periodic Snapshots Designs 258
  - Transactions 258
  - Snapshots 259
  - Semi-Additivity 260
  - Invisible Aggregates for Periodic Snapshots 261
  - Averaging Semi-Additive Facts Produces a Derived Schema 263
  - Taking Less Frequent Snapshots Does Not Produce an Invisible Aggregate 264

Accumulating Snapshots 265
- The Accumulating Snapshot 265
- Aggregating the Accumulating Snapshot 267

Factless Fact Tables 269
- Factless Events and Aggregates 269
- Coverage Tables and Aggregates 271
Aggregating Dimensions 272
Transaction Dimensions 273
Timestamping a Dimension 273
Aggregating a Timestamped Dimension 274
Bridge Tables 275
Dealing with Multi-Valued Attributes 276
Aggregates and Bridge Tables 278
Core and Custom Stars 282
Other Schema Types 283
Snowflakes and Aggregates 283
The Snowflake Schema 283
Aggregating Snowflakes 284
Third Normal Form Schemas and Aggregates 287
Summary 290

Chapter 9  Related Topics 291
Aggregates and the Archive Strategy 292
The Data Warehouse Archive Strategy 292
Aggregates and Archives 295
Maintaining Aggregates 295
Archive Versus Purge 297
Summarizing Off-Line Data 297
Aggregates and Security 299
Dimensionally Driven Security and Aggregates 299
Unrestricted Access to Summary Data 301
Derived Tables 302
The Merged Fact Table 303
The Pivoted Fact Table 307
The Sliced Fact Table 309
When Rollups Are Deployed Before Detail 310
Building the Base Table First 311
Building the Rollup First 312
Parallel Load Processes 312
Redeveloping the Load 314
Historic Detail 316
Summary 317

Glossary 319
Index 329