

- Acknowledgments p. xvii
- About the Authors p. xix
- Introduction p. xxi
- Part I Requirements, Realities, and Architecture p. 1
- Chapter 1 Surrounding the Requirements p. 3
 - Requirements p. 4
 - Architecture p. 9
 - The Mission of the Data Warehouse p. 22
 - The Mission of the ETL Team p. 28
- Chapter 2 ETL Data Structures p. 29
 - To Stage or Not to Stage p. 29
 - Designing the Staging Area p. 31
 - Data Structures in the ETL System p. 35
 - Planning and Design Standards p. 48
 - Summary p. 52
- Part II Data Flow p. 53
- Chapter 3 Extracting p. 55
 - Part 1 The Logical Data Map p. 56
 - Inside the Logical Data Map p. 58
 - Building the Logical Data Map p. 62
 - Integrating Heterogeneous Data Sources p. 73
 - Part 2 The Challenge of Extracting from Disparate Platforms p. 76
 - Mainframe Sources p. 78
 - Flat Files p. 90
 - XML Sources p. 93
 - Web Log Sources p. 97
 - ERP System Sources p. 102
 - Part 3 Extracting Changed Data p. 105
 - Summary p. 111
- Chapter 4 Cleaning and Conforming p. 113
 - Defining Data Quality p. 115
 - Assumptions p. 116
 - Part 1 Design Objectives p. 117
 - Part 2 Cleaning Deliverables p. 124
 - Part 3 Screens and Their Measurements p. 131
 - Part 4 Conforming Deliverables p. 148
 - Summary p. 160
- Chapter 5 Delivering Dimension Tables p. 161
 - The Basic Structure of a Dimension p. 162
 - The Grain of a Dimension p. 165
 - The Basic Load Plan for a Dimension p. 166
 - Flat Dimensions and Snowflaked Dimensions p. 167
 - Date and Time Dimensions p. 170
 - Big Dimensions p. 174

- Small Dimensions p. 176
- One Dimension or Two p. 176
- Dimensional Roles p. 178
- Dimensions as Subdimensions of Another Dimension p. 180
- Degenerate Dimensions p. 182
- Slowly Changing Dimensions p. 183
- Type 1 Slowly Changing Dimension (Overwrite) p. 183
- Type 2 Slowly Changing Dimension (Partitioning History) p. 185
- Late-Arriving Dimension Records and Correcting Bad Data p. 194
- Precise Time Stamping of a Type 2 Slowly Changing Dimension p. 190
- Type 3 Slowly Changing Dimension (Alternate Realities) p. 192
- Hybrid Slowly Changing Dimensions p. 193
- Multivalued Dimensions and Bridge Tables p. 196
- Ragged Hierarchies and Bridge Tables p. 199
- Technical Note: Populating Hierarchy Bridge Tables p. 201
- Using Positional Attributes in a Dimension to Represent Text Facts p. 204
- Summary p. 207
- Chapter 6 Delivering Fact Tables p. 209
- The Basic Structure of a Fact Table p. 210
- Guaranteeing Referential Integrity p. 212
- Surrogate Key Pipeline p. 214
- Fundamental Grains p. 217
- Preparing for Loading Fact Tables p. 224
- Factless Fact Tables p. 232
- Augmenting a Type 1 Fact Table with Type 2 History p. 234
- Graceful Modifications p. 235
- Multiple Units of Measure in a Fact Table p. 237
- Collecting Revenue in Multiple Currencies p. 238
- Late Arriving Facts p. 239
- Aggregations p. 241
- Delivering Dimensional Data to OLAP Cubes p. 247
- Summary p. 253
- Part III Implementation and operations p. 255
- Chapter 7 Development p. 257
- Current Marketplace ETL Tool Suite Offerings p. 258
- Current Scripting Languages p. 260
- Time Is of the Essence p. 260
- Using Database Bulk Loader Utilities to Speed Inserts p. 276
- Managing Database Features to Improve Performance p. 280
- Troubleshooting Performance Problems p. 292
- Increasing ETL Throughput p. 294
- Summary p. 300
- Chapter 8 Operations p. 301
- Scheduling and Support p. 302

- Migrating to Production p. 315
- Achieving Optimal ETL Performance p. 320
- Purging Historic Data p. 330
- Monitoring the ETL System p. 331
- Tuning ETL Processes p. 339
- ETL System Security p. 343
- Short-Term Archiving and Recovery p. 345
- Long-Term Archiving and Recovery p. 346
- Summary p. 350
- Chapter 9 Metadata p. 351
- Defining Metadata p. 352
- Business Metadata p. 359
- Technical Metadata p. 363
- ETL-Generated Metadata p. 367
- Metadata Standards and Practices p. 377
- Impact Analysis p. 380
- Summary p. 380
- Chapter 10 Responsibilities p. 383
- Planning and Leadership p. 383
- Managing the Project p. 391
- Summary p. 416
- Part IV Real Time Streaming ETL Systems p. 419
- Chapter 11 Real-Time ETL Systems p. 421
- Why Real-Time ETL? p. 422
- Defining Real-Time ETL p. 424
- Challenges and Opportunities of Real-Time Data Warehousing p. 424
- Real-Time Data Warehousing Review p. 425
- Categorizing the Requirement p. 430
- Real-Time ETL Approaches p. 437
- Summary p. 459
- Chapter 12 Conclusions p. 461
- Deepening the Definition of ETL p. 461
- The Future of Data Warehousing and ETL in Particular p. 463
- Index p. 467