Building a Data Warehouse

With Examples in SQL Server

- EiD

Vincent Rainardi

Apress®

Contents

About the Author xiii
Preface xv

^CHAPTER 1 Introduction to Data Warehousing 1

What Is a Data Warehouse? 1
Retrieves Data 4
Consolidates Data 5
Periodically 6
Dimensional Data Store 7
Normalized Data Store 8
History 10
Query 11
Business Intelligence 12
Other Analytical Activities 14
Updated in Batches 15
Other Definitions 16
Data Warehousing Today 17
Business Intelligence 17
Customer Relationship Management 18
Data Mining 19
Master Data Management (MDM) 20
Customer Data Integration 23
Future Trends in Data Warehousing 24
Unstructured Data 24
Search 25
Service-Oriented Architecture (SOA) 26
Real-Time Data Warehouse 27
Summary 27

1CHAPTER 2 Data Warehouse Architecture 29

Data Flow Architecture 29
Single DDS 33
NDS + DDS 35
^ODS + DDS 38
Federated Data Warehouse 39
System Architecture 42
Case Study 44
Summary 47

•CHAPTER 3 Data Warehouse Development Methodology 49

Waterfall Methodology 49
Iterative Methodology 54
Summary 59

•CHAPTER 4 Functional and Nonfunctional Requirements 61

Identifying Business Areas 61
Understanding Business Operations 62
Defining Functional Requirements 63
Defining Nonfunctional Requirements 65
Conducting a Data Feasibility Study 67
Summary 70

KHAPTER 5 Data Modeling 71

Designing the Dimensional Data Store 71
Dimension Tables 76
Date Dimension 77
Slowly Changing Dimension 80
Product, Customer, and Store Dimensions 83
Subscription Sales Data Mart 89
Supplier Performance Data Mart 94
CRM Data Marts^ 96
Data Hierarchy 101
Source System Mapping 102
Designing the Normalized Data Store 106
Summary 111

...CHAPTER 6 Physical Database Design ... 113

Hardware Platform ... 113
Storage Considerations ... 120
Configuring Databases ... 123
Creating DDS Database Structure ... 128
Creating the Normalized Data Store ... 139
Using Views ... 157
Summary Tables ... 161
Partitioning ... 162
Indexes ... 166
Summary ... 171

SCHAPTER 7 Data Extraction ... 173

Introduction to ETL ... 173
ETL Approaches and Architecture ... 174
General Considerations ... 177
Extracting Relational Databases ... 180
Whole Table Every Time ... 180
Incremental Extract ... 181
Fixed Range ... 185
Related Tables ... 186
Testing Data Leaks ... 187
Extracting File Systems ... 187
Extracting Other Source Types ... 190
Extracting Data Using SSIS ... 191
Memorizing the Last Extraction Timestamp ... 200
Extracting from Files ... 208
Summary ... 214

•CHAPTER 8 Populating the Data Warehouse ... 215

Stage Loading ... 216
Data Firewall ... 218
Populating NDS ... 219
Using SSIS to Populate NDS ... 228
Upsert Using SQL and Lookup ... 235
Normalization ... 242
Practical Tips on SSIS ... 249

Populating DDS Dimension Tables . . . 250
Populating DDS Fact Tables . . . 266
Batches, Mini-batches, and Near Real-Time ETL . . . 269
Pushing the Data In . . . 270
Summary . . . 271

ICHAPTER 9 Assuring Data Quality . . . 273

Data Quality Process . . . 274
Data Cleansing and Matching . . . 277
Cross-checking with External Sources . . . 290
Data Quality Rules . . . 291
Action: Reject, Allow, Fix . . . 293
Logging and Auditing . . . 296
Data Quality Reports and Notifications . . . 298
Summary . . . 300

ICHAPTER 10 Metadata . . . 301

Metadata in Data Warehousing . . . 301
Data Definition and Mapping Metadata . . . 303
Data Structure Metadata . . . 308
Source System Metadata . . . 313
ETL Process Metadata . . . 318
Data Quality Metadata . . . 320
Audit Metadata . . . 323
Usage Metadata . . . 324
Maintaining Metadata . . . 325
Summary . . . 327

1CHAPTER 11 Building Reports . . . 329

Data Warehouse Reports . . . 329
When to Use Reports and When Not to Use Them . . . 332
Report Wizard . . . 334
Report Layout . . . 340
Report Parameters . . . 342
Grouping, Sorting, and Filtering . . . 351
Simplicity . . . 356
Spreadsheets . . . 357
Multidimensional Database Reports . . . 362
Deploying Reports . . . 366

Managing Reports 370
Managing Report Security 370
Managing Report Subscriptions 372
Managing Report Execution 374
Summary 375

[CHAPTER 12 Multidimensional Database 377

What a Multidimensional Database Is 377
Online Analytical Processing 380
Creating a Multidimensional Database 381
Processing a Multidimensional Database 388
Querying a Multidimensional Database 394
Administering a Multidimensional Database 396
Multidimensional Database Security 397
Processing Cubes 399
Backup and Restore 405
Summary 409

ICHAPTER 13 Using Data Warehouse for Business Intelligence 411

Business Intelligence Reports 412
Business Intelligence Analytics 413
Business Intelligence Data Mining 416
Business Intelligence Dashboards 432
Business Intelligence Alerts 437
Business Intelligence Portal 438
Summary 439

CHAPTER 14 Using Data Warehouse for Customer Relationship Management 441

Single Customer View 442
Campaign Segmentation 447
Permission Management 450
Delivery and Response Data 454
Customer Analysis 460
Customer Support 463
Personalization 464
Customer Loyalty Scheme 465
Summary 466

ICHAPTER 15 Other Data Warehouse Usage 467

Customer Data Integration 467
Unstructured Data 470
Search in Data Warehousing 474
Summary 476

ICHAPTER 16 Testing Your Data Warehouse 477

Data Warehouse ETL Testing 478
Functional Testing 480
Performance Testing 482
Security Testing 485
User Acceptance Testing 486
End-to-End Testing 487
Migrating to Production 487
Summary 489

ICHAPTER 17 Data Warehouse Administration 491

Monitoring Data Warehouse ETL 492
Monitoring Data Quality 495
Managing Security 498
Managing Databases 499
Making Schema Changes 501
Updating Applications 503
Summary 503

APPENDIX Normalization Rules 505

IINDEX 509