

Contents

Part I Fundamental Concepts

1	Introduction	3
1.1	An Overview of Data Warehousing	4
1.2	Emerging Data Warehousing Technologies	7
1.3	Review Questions	10
2	Database Concepts	11
2.1	Database Design	11
2.2	The Northwind Case Study	13
2.3	Conceptual Database Design	13
2.4	Logical Database Design	18
2.4.1	The Relational Model	18
2.4.2	Normalization	24
2.4.3	Relational Query Languages	26
2.5	Physical Database Design	36
2.6	Summary	39
2.7	Bibliographic Notes	40
2.8	Review Questions	40
2.9	Exercises	41
3	Data Warehouse Concepts	45
3.1	Multidimensional Model	45
3.1.1	Hierarchies	47
3.1.2	Measures	49
3.2	OLAP Operations	51
3.3	Data Warehouses	63
3.4	Data Warehouse Architecture	67
3.4.1	Back-End Tier	67
3.4.2	Data Warehouse Tier	68
3.4.3	OLAP Tier	69

3.4.4	Front-End Tier	70
3.4.5	Variations of the Architecture	70
3.5	Overview of Microsoft SQL Server BI Tools	71
3.6	Summary	72
3.7	Bibliographic Notes	72
3.8	Review Questions	73
3.9	Exercises	73
4	Conceptual Data Warehouse Design	75
4.1	Conceptual Modeling of Data Warehouses	75
4.2	Hierarchies	79
4.2.1	Balanced Hierarchies	79
4.2.2	Unbalanced Hierarchies	80
4.2.3	Generalized Hierarchies	81
4.2.4	Alternative Hierarchies	83
4.2.5	Parallel Hierarchies	84
4.2.6	Nonstrict Hierarchies	87
4.3	Advanced Modeling Aspects	90
4.3.1	Facts with Multiple Granularities	91
4.3.2	Many-to-Many Dimensions	91
4.3.3	Links between Facts	95
4.4	Querying the Northwind Cube Using the OLAP Operations ..	96
4.5	Summary	99
4.6	Bibliographic Notes	100
4.7	Review Questions	101
4.8	Exercises	102
5	Logical Data Warehouse Design	105
5.1	Logical Modeling of Data Warehouses	105
5.2	Relational Data Warehouse Design	106
5.3	Relational Representation of Data Warehouses	109
5.4	Time Dimension	112
5.5	Logical Representation of Hierarchies	113
5.5.1	Balanced Hierarchies	113
5.5.2	Unbalanced Hierarchies	114
5.5.3	Generalized Hierarchies	115
5.5.4	Alternative Hierarchies	117
5.5.5	Parallel Hierarchies	118
5.5.6	Nonstrict Hierarchies	119
5.6	Advanced Modeling Aspects	120
5.6.1	Facts with Multiple Granularities	120
5.6.2	Many-to-Many Dimensions	121
5.6.3	Links between Facts	122
5.7	Slowly Changing Dimensions	124
5.8	Performing OLAP Queries with SQL	130

- 5.9 Defining the Northwind Data Warehouse in Analysis Services 135
 - 5.9.1 Multidimensional Model 135
 - 5.9.2 Tabular Model 147
- 5.10 Summary 155
- 5.11 Bibliographic Notes 156
- 5.12 Review Questions 156
- 5.13 Exercises 157
- 6 Data Analysis in Data Warehouses 159**
 - 6.1 Introduction to MDX 160
 - 6.1.1 Tuples and Sets 160
 - 6.1.2 Basic Queries 162
 - 6.1.3 Slicing 163
 - 6.1.4 Navigation 164
 - 6.1.5 Cross Join 166
 - 6.1.6 Subqueries 167
 - 6.1.7 Calculated Members and Named Sets 168
 - 6.1.8 Relative Navigation 170
 - 6.1.9 Time-Related Calculations 171
 - 6.1.10 Filtering 174
 - 6.1.11 Sorting 175
 - 6.1.12 Top and Bottom Analysis 176
 - 6.1.13 Aggregation Functions 177
 - 6.2 Introduction to DAX 179
 - 6.2.1 Expressions 179
 - 6.2.2 Evaluation Context 181
 - 6.2.3 Queries 183
 - 6.2.4 Filtering 184
 - 6.2.5 Hierarchy Handling 188
 - 6.2.6 Time-Related Calculations 189
 - 6.2.7 Top and Bottom Analysis 192
 - 6.2.8 Table Operations 194
 - 6.3 Key Performance Indicators 196
 - 6.3.1 Classification of Key Performance Indicators 197
 - 6.3.2 Defining Key Performance Indicators 198
 - 6.4 Dashboards 200
 - 6.4.1 Types of Dashboards 201
 - 6.4.2 Guidelines for Dashboard Design 202
 - 6.5 Summary 203
 - 6.6 Bibliographic Notes 203
 - 6.7 Review Questions 204

7 Data Analysis in the Northwind Data Warehouse 205

7.1 Querying the Multidimensional Model in MDX 205

7.2 Querying the Tabular Model in DAX 211

7.3 Querying the Relational Data Warehouse in SQL 217

7.4 Comparison of MDX, DAX, and SQL 225

7.5 KPIs for the Northwind Case Study 229

7.5.1 KPIs in Analysis Services Multidimensional 229

7.5.2 KPIs in Analysis Services Tabular 232

7.6 Dashboards for the Northwind Case Study 234

7.6.1 Dashboards in Reporting Services 235

7.6.2 Dashboards in Power BI 239

7.7 Summary 241

7.8 Review Questions 241

7.9 Exercises 242

Part II Implementation and Deployment

8 Physical Data Warehouse Design 245

8.1 Physical Modeling of Data Warehouses 246

8.2 Materialized Views 247

8.2.1 Algorithms Using Full Information 249

8.2.2 Algorithms Using Partial Information 251

8.3 Data Cube Maintenance 252

8.4 Computation of a Data Cube 258

8.4.1 PipeSort Algorithm 259

8.4.2 Cube Size Estimation 262

8.4.3 Partial Computation of a Data Cube 263

8.5 Indexes for Data Warehouses 267

8.5.1 Bitmap Indexes 268

8.5.2 Bitmap Compression 271

8.5.3 Join Indexes 272

8.6 Evaluation of Star Queries 273

8.7 Partitioning 275

8.8 Parallel Processing 277

8.9 Physical Design in SQL Server and Analysis Services 280

8.9.1 Indexed Views 280

8.9.2 Partition-Aligned Indexed Views 282

8.9.3 Column-Store Indexes 283

8.9.4 Partitions in Analysis Services 284

8.10 Query Performance in Analysis Services 286

8.11 Summary 289

8.12 Bibliographic Notes 290

8.13 Review Questions 290

8.14 Exercises 291

9 Extraction, Transformation, and Loading 297

9.1 Business Process Modeling Notation 298

9.2 Conceptual ETL Design Using BPMN 303

9.3 Conceptual Design of the Northwind ETL Process 306

9.4 SQL Server Integration Services 318

9.5 The Northwind ETL Process in Integration Services..... 320

9.6 Implementing ETL Processes in SQL 326

9.7 Summary 332

9.8 Bibliographic Notes 332

9.9 Review Questions 333

9.10 Exercises 334

10 A Method for Data Warehouse Design..... 335

10.1 Approaches to Data Warehouse Design 335

10.2 General Overview of the Method 337

10.3 Requirements Specification..... 338

10.3.1 Business-Driven Requirements Specification 339

10.3.2 Data-driven Requirements Specification..... 345

10.3.3 Business/Data-driven Requirements Specification ... 349

10.4 Conceptual Design 350

10.4.1 Business-Driven Conceptual Design 351

10.4.2 Data-driven Conceptual Design 354

10.4.3 Business/Data-driven Conceptual Design 356

10.5 Logical Design 357

10.5.1 Logical Schemas 358

10.5.2 ETL Processes 359

10.6 Physical Design 359

10.7 Characterization of the Various Approaches 360

10.7.1 Business-Driven Approach 360

10.7.2 Data-driven Approach 361

10.7.3 Business/Data-driven Approach 362

10.8 Summary 363

10.9 Bibliographic Notes 363

10.10 Review Questions..... 365

10.11 Exercises 366

Part III Advanced Topics

11 Temporal and Multiversion Data Warehouses..... 373

11.1 Manipulating Temporal Information in SQL..... 374

11.2 Conceptual Design of Temporal Data Warehouses..... 383

11.2.1 Time Data Types 383

11.2.2 Synchronization Relationships 384

11.2.3 A Conceptual Model for Temporal Data Warehouses 386

11.2.4 Temporal Hierarchies 389

11.2.5	Temporal Facts	391
11.3	Logical Design of Temporal Data Warehouses	392
11.4	Implementation Considerations	395
11.4.1	Period Encoding	395
11.4.2	Tables for Temporal Roll-Up	395
11.4.3	Integrity Constraints	396
11.4.4	Measure Aggregation	399
11.4.5	Temporal Measures	403
11.5	Querying the Temporal Northwind Data Warehouse in SQL ..	404
11.6	Temporal Data Warehouses versus Slowly Changing Dimensions	412
11.7	Conceptual Design of Multiversion Data Warehouses	416
11.8	Logical Design of Multiversion Data Warehouses	422
11.9	Querying the Multiversion Northwind Data Warehouse in SQL	427
11.10	Summary	428
11.11	Bibliographic Notes	429
11.12	Review Questions	430
11.13	Exercises	431
12	Spatial and Mobility Data Warehouses	437
12.1	Conceptual Design of Spatial Data Warehouses	438
12.1.1	Spatial Data Types	438
12.1.2	Topological relationships	440
12.1.3	Continuous Fields	441
12.1.4	A Conceptual Model of Spatial Data Warehouses ...	441
12.2	Implementation Considerations for Spatial Data	445
12.2.1	Spatial Reference Systems	445
12.2.2	Vector Model	447
12.2.3	Raster Model	449
12.3	Logical Design of Spatial Data Warehouses	451
12.4	Topological Constraints	454
12.5	Querying the GeoNorthwind Data Warehouse in SQL	456
12.6	Mobility Data Analysis	460
12.7	Temporal Types	461
12.8	Temporal Types in MobilityDB	466
12.9	Mobility Data Warehouses	470
12.10	Querying the Northwind Mobility Data Warehouse in SQL ..	474
12.11	Summary	480
12.12	Bibliographic Notes	480
12.13	Review Questions	481
12.14	Exercises	482

- 13 Graph Data Warehouses** 487
 - 13.1 Graph Data Models 488
 - 13.2 Property Graph Database Systems 490
 - 13.2.1 Neo4j 492
 - 13.2.2 Introduction to Cypher 493
 - 13.2.3 Querying the Northwind Cube with Cypher 501
 - 13.3 OLAP on Hypergraphs 507
 - 13.3.1 Operations on Hypergraphs 512
 - 13.3.2 OLAP on Trajectory Graphs 516
 - 13.4 Graph Processing Frameworks 520
 - 13.4.1 Gremlin 520
 - 13.4.2 JanusGraph 523
 - 13.5 Bibliographic Notes 526
 - 13.6 Review Questions 526
 - 13.7 Exercises 527
- 14 Semantic Web Data Warehouses** 531
 - 14.1 Semantic Web 532
 - 14.1.1 Introduction to RDF and RDFS 532
 - 14.1.2 RDF Serializations 533
 - 14.1.3 RDF Representation of Relational Data 535
 - 14.2 Introduction to SPARQL 539
 - 14.2.1 SPARQL Basics 540
 - 14.2.2 SPARQL Semantics 543
 - 14.3 RDF Representation of Multidimensional Data 544
 - 14.4 Representation of the Northwind Cube in QB4OLAP 547
 - 14.5 Querying the Northwind Cube in SPARQL 549
 - 14.6 Summary 557
 - 14.7 Bibliographic Notes 557
 - 14.8 Review Questions 558
 - 14.9 Exercises 559
- 15 Recent Developments in Big Data Warehouses** 561
 - 15.1 Data Warehousing in the Age of Big Data 562
 - 15.2 Distributed Processing Frameworks 563
 - 15.2.1 Hadoop 565
 - 15.2.2 Hive 567
 - 15.2.3 Spark 569
 - 15.2.4 Comparison of Hadoop and Spark 576
 - 15.2.5 Kylin 577
 - 15.3 Distributed Database Systems 579
 - 15.3.1 MySQL Cluster 582
 - 15.3.2 Citus 585
 - 15.4 In-Memory Database Systems 587
 - 15.4.1 Oracle TimesTen 590

15.4.2	Redis	591
15.5	Column-Store Database Systems	592
15.5.1	Vertica	595
15.5.2	MonetDB	597
15.5.3	Citus Columnar	598
15.6	NoSQL Database Systems	599
15.6.1	HBase	600
15.6.2	Cassandra	602
15.7	NewSQL Database Systems	606
15.7.1	Cloud Spanner	607
15.7.2	SAP HANA	607
15.7.3	VoltDB	609
15.8	Array Database Systems	610
15.8.1	Rasdaman	612
15.8.2	SciDB	614
15.9	Hybrid Transactional and Analytical Processing	616
15.9.1	SingleStore	617
15.9.2	LeanXcale	618
15.10	Polystores	619
15.10.1	CloudMdsQL	620
15.10.2	BigDAWG	621
15.11	Cloud Data Warehouses	622
15.12	Data Lakes and Data Lakehouses	624
15.13	Future Perspectives	628
15.14	Summary	629
15.15	Bibliographic Notes	629
15.16	Review Questions	630
A	Graphical Notation	633
A.1	Entity-Relationship Model	633
A.2	Relational Model	635
A.3	MultiDim Model for Data Warehouses	635
A.4	MultiDim Model for Spatial Data Warehouses	639
A.5	MultiDim Model for Temporal Data Warehouses	641
A.6	BPMN Notation for ETL	643
References		647
Glossary		667
Index		685