

Part I Fundamental Concepts

1	Introduction	3
1.1	A Historical Overview of Data Warehousing	4
1.2	Spatial and Spatiotemporal Data Warehouses	8
1.3	New Domains and Challenges	9
1.4	Review Questions	11
2	Database Concepts	13
2.1	Database Design	13
2.2	The Northwind Case Study	15
2.3	Conceptual Database Design	16
2.4	Logical Database Design	21
2.4.1	The Relational Model	21
2.4.2	Normalization	27
2.4.3	Relational Query Languages	30
2.5	Physical Database Design	43
2.6	Summary	46
2.7	Bibliographic Notes	47
2.8	Review Questions	47
2.9	Exercises	48
3	Data Warehouse Concepts	53
3.1	Multidimensional Model	53
3.1.1	Hierarchies	56
3.1.2	Measures	57
3.2	OLAP Operations	59
3.3	Data Warehouses	72
3.4	Data Warehouse Architecture	76
3.4.1	Back-End Tier	76
3.4.2	Data Warehouse Tier	77
3.4.3	OLAP Tier	78

3.4.4	Front-End Tier	79
3.4.5	Variations of the Architecture	79
3.5	Data Warehouse Design	80
3.6	Business Intelligence Tools	81
3.6.1	Overview of Microsoft SQL Server Tools	82
3.6.2	Overview of Pentaho Business Analytics	83
3.7	Summary	84
3.8	Bibliographic Notes	84
3.9	Review Questions	85
3.10	Exercises	86
4	Conceptual Data Warehouse Design	89
4.1	Conceptual Modeling of Data Warehouses	89
4.2	Hierarchies	94
4.2.1	Balanced Hierarchies	95
4.2.2	Unbalanced Hierarchies	95
4.2.3	Generalized Hierarchies	96
4.2.4	Alternative Hierarchies	98
4.2.5	Parallel Hierarchies	99
4.2.6	Nonstrict Hierarchies	102
4.3	Advanced Modeling Aspects	106
4.3.1	Facts with Multiple Granularities	106
4.3.2	Many-to-Many Dimensions	106
4.4	Querying the Northwind Cube Using the OLAP Operations	110
4.5	Summary	114
4.6	Bibliographic Notes	115
4.7	Review Questions	116
4.8	Exercises	116
5	Logical Data Warehouse Design	121
5.1	Logical Modeling of Data Warehouses	121
5.2	Relational Data Warehouse Design	123
5.3	Relational Implementation of the Conceptual Model	126
5.4	Time Dimension	128
5.5	Logical Representation of Hierarchies	129
5.5.1	Balanced Hierarchies	129
5.5.2	Unbalanced Hierarchies	130
5.5.3	Generalized Hierarchies	132
5.5.4	Alternative Hierarchies	134
5.5.5	Parallel Hierarchies	134
5.5.6	Nonstrict Hierarchies	135
5.6	Advanced Modeling Aspects	136
5.6.1	Facts with Multiple Granularities	137
5.6.2	Many-to-Many Dimensions	138
5.7	Slowly Changing Dimensions	139

5.8	SQL/OLAP Operations	145
5.8.1	Data Cube	146
5.8.2	ROLLUP, CUBE, and GROUPING SETS	147
5.8.3	Window Functions	149
5.9	Definition of the Northwind Cube in Analysis Services	152
5.9.1	Data Sources	152
5.9.2	Data Source Views	152
5.9.3	Dimensions	154
5.9.4	Hierarchies	158
5.9.5	Cubes	161
5.10	Definition of the Northwind Cube in Mondrian	164
5.10.1	Schemas and Physical Schemas	165
5.10.2	Cubes, Dimensions, Attributes, and Hierarchies ...	166
5.10.3	Measures	171
5.11	Summary	173
5.12	Bibliographic Notes	173
5.13	Review Questions	173
5.14	Exercises	174
6	Querying Data Warehouses	179
6.1	Introduction to MDX	180
6.1.1	Tuples and Sets	180
6.1.2	Basic Queries	181
6.1.3	Slicing	183
6.1.4	Navigation	185
6.1.5	Cross Join	188
6.1.6	Subqueries	189
6.1.7	Calculated Members and Named Sets	191
6.1.8	Relative Navigation	193
6.1.9	Time Series Functions	196
6.1.10	Filtering	200
6.1.11	Sorting	201
6.1.12	Top and Bottom Analysis	203
6.1.13	Aggregation Functions	205
6.2	Querying the Northwind Cube in MDX	207
6.3	Querying the Northwind Data Warehouse in SQL	216
6.4	Comparison of MDX and SQL	225
6.5	Summary	227
6.6	Bibliographic Notes	228
6.7	Review Questions	230
6.8	Exercises	230

Part II Implementation and Deployment

7	Physical Data Warehouse Design	233
7.1	Physical Modeling of Data Warehouses	234
7.2	Materialized Views	235
7.2.1	Algorithms Using Full Information	237
7.2.2	Algorithms Using Partial Information	239
7.3	Data Cube Maintenance	240
7.4	Computation of a Data Cube	246
7.4.1	PipeSort Algorithm	247
7.4.2	Cube Size Estimation	250
7.4.3	Partial Computation of a Data Cube	251
7.5	Indexes for Data Warehouses	256
7.5.1	Bitmap Indexes	257
7.5.2	Bitmap Compression	259
7.5.3	Join Indexes	260
7.6	Evaluation of Star Queries	261
7.7	Data Warehouse Partitioning	263
7.7.1	Queries in Partitioned Databases	264
7.7.2	Managing Partitioned Databases	265
7.7.3	Partitioning Strategies	265
7.8	Physical Design in SQL Server and Analysis Services	266
7.8.1	Indexed Views	266
7.8.2	Partition-Aligned Indexed Views	267
7.8.3	Column-Store Indexes	269
7.8.4	Partitions in Analysis Services	269
7.9	Query Performance in Analysis Services	274
7.10	Query Performance in Mondrian	276
7.10.1	Aggregate Tables	276
7.10.2	Caching	277
7.11	Summary	278
7.12	Bibliographic Notes	279
7.13	Review Questions	279
7.14	Exercises	280
8	Extraction, Transformation, and Loading	285
8.1	Business Process Modeling Notation	286
8.2	Conceptual ETL Design Using BPMN	291
8.3	Conceptual Design of the Northwind ETL Process	295
8.4	Integration Services and Kettle	309
8.4.1	Overview of Integration Services	309
8.4.2	Overview of Kettle	311
8.5	The Northwind ETL Process in Integration Services	312
8.6	The Northwind ETL Process in Kettle	319
8.7	Summary	324

8.8	Bibliographic Notes	325
8.9	Review Questions.....	325
8.10	Exercises	326
9	Data Analytics: Exploiting the Data Warehouse	329
9.1	Data Mining	330
9.1.1	Data Mining Tasks	331
9.1.2	Supervised Classification	333
9.1.3	Clustering	336
9.1.4	Association Rules	338
9.1.5	Pattern Growth Algorithm	344
9.1.6	Sequential Patterns	347
9.1.7	Data Mining in Analysis Services	350
9.2	Key Performance Indicators	362
9.2.1	Classification of Key Performance Indicators	363
9.2.2	Guidelines for Defining Key Performance Indicators	364
9.2.3	KPIs for the Northwind Case Study	366
9.2.4	KPIs in Analysis Services.....	367
9.3	Dashboards	370
9.3.1	Types of Dashboards	371
9.3.2	Guidelines for Dashboard Design	372
9.3.3	Dashboards in Reporting Services	373
9.4	Summary	378
9.5	Bibliographic Notes	378
9.6	Review Questions.....	379
9.7	Exercises	380
10	A Method for Data Warehouse Design	385
10.1	Approaches to Data Warehouse Design.....	386
10.2	General Overview of the Method.....	388
10.3	Requirements Specification	389
10.3.1	Analysis-Driven Requirements Specification	389
10.3.2	Analysis-Driven Requirements for the Northwind Case Study	392
10.3.3	Source-Driven Requirements Specification	396
10.3.4	Source-Driven Requirements for the Northwind Case Study	398
10.3.5	Analysis/Source-Driven Requirements Specification	401
10.4	Conceptual Design	402
10.4.1	Analysis-Driven Conceptual Design.....	402
10.4.2	Analysis-Driven Conceptual Design for the Northwind Case Study	404
10.4.3	Source-Driven Conceptual Design.....	407

10.4.4	Source-Driven Conceptual Design for the Northwind Case Study	408
10.4.5	Analysis/Source-Driven Conceptual Design	409
10.5	Logical Design	410
10.5.1	Logical Schemas	411
10.5.2	ETL Processes	413
10.6	Physical Design	413
10.7	Characterization of the Various Approaches	415
10.7.1	Analysis-Driven Approach	415
10.7.2	Source-Driven Approach	416
10.7.3	Analysis/Source-Driven Approach	417
10.8	Summary	418
10.9	Bibliographic Notes	418
10.10	Review Questions	419
10.11	Exercises	420

Part III Advanced Topics

11	Spatial Data Warehouses	427
11.1	General Concepts of Spatial Databases	428
11.1.1	Spatial Data Types	428
11.1.2	Continuous Fields	432
11.2	Conceptual Modeling of Spatial Data Warehouses	434
11.2.1	Spatial Hierarchies	438
11.2.2	Spatiality and Measures	440
11.3	Implementation Considerations for Spatial Data	442
11.3.1	Spatial Reference Systems	442
11.3.2	Vector Model	443
11.3.3	Raster Model	446
11.4	Relational Representation of Spatial Data Warehouses	448
11.4.1	Spatial Levels and Attributes	448
11.4.2	Spatial Facts, Measures, and Hierarchies	450
11.4.3	Topological Constraints	452
11.5	GeoMondrian	454
11.6	Querying the GeoNorthwind Cube in MDX	455
11.7	Querying the GeoNorthwind Data Warehouse in SQL	459
11.8	Spatial Data Warehouse Design	461
11.8.1	Requirements Specification and Conceptual Design	462
11.8.2	Logical and Physical Design	467
11.9	Summary	467
11.10	Bibliographic Notes	468
11.11	Review Questions	468
11.12	Exercises	469

12	Trajectory Data Warehouses	475
12.1	Mobility Data Analysis	476
12.2	Temporal Types	477
12.2.1	Temporal Spatial Types	481
12.2.2	Temporal Field Types	483
12.3	Implementation of Temporal Types in PostGIS	485
12.4	The Northwind Trajectory Data Warehouse	490
12.5	Querying the Northwind Trajectory Data Warehouse in SQL	495
12.6	Summary	502
12.7	Bibliographic Notes	502
12.8	Review Questions	503
12.9	Exercises	504
13	New Data Warehouse Technologies	507
13.1	MapReduce and Hadoop	508
13.2	High-Level Languages for Hadoop	510
13.2.1	Hive	510
13.2.2	Pig Latin	512
13.3	Column-Store Database Systems	514
13.4	In-Memory Database Systems	516
13.5	Representative Systems	519
13.5.1	Vertica	519
13.5.2	MonetDB	520
13.5.3	MonetDB/X100	521
13.5.4	SAP HANA	522
13.5.5	Oracle TimesTen	524
13.5.6	SQL Server xVelocity	526
13.6	Real-Time Data Warehouses	528
13.7	Extraction, Loading, and Transformation	532
13.8	Summary	534
13.9	Bibliographic Notes	535
13.10	Review Questions	535
13.11	Exercises	536
14	Data Warehouses and the Semantic Web	539
14.1	Semantic Web	540
14.1.1	Introduction to RDF and RDFS	540
14.1.2	RDF Serializations	541
14.1.3	RDF Representation of Relational Data	543
14.2	SPARQL	547
14.3	RDF Representation of Multidimensional Data	551
14.3.1	RDF Data Cube Vocabulary	553
14.3.2	QB4OLAP Vocabulary	557
14.4	Representation of the Northwind Cube in QB4OLAP	561
14.5	Querying the Northwind Cube in SPARQL	564

14.6	Summary	573
14.7	Bibliographic Notes	574
14.8	Review Questions.....	575
14.9	Exercises	575
15	Conclusion	577
15.1	Temporal Data Warehouses	577
15.2	3D/4D Spatial Data Warehouses	579
15.3	Text Analytics and Text Data Warehouses.....	581
15.4	Multimedia Data Warehouses	583
15.5	Graph Analytics and Graph Data Warehouses	586
A	Graphical Notation	589
A.1	Entity-Relationship Model.....	589
A.2	Relational Model	591
A.3	MultiDim Model for Data Warehouses	591
A.4	MultiDim Model for Spatial Data Warehouses.....	595
A.5	BPMN Notation for ETL	597
	References.....	601
	Index	615