



Veit
Köppen

Gunter
Saake

Kai-Uwe
Sattler

2. Auflage

Data Warehouse Technologien

Inhaltsverzeichnis

Inhaltsverzeichnis	ix
1 Einführung in Data-Warehouse-Systeme	1
1.1 Anwendungsszenario Getränkemarkt	2
1.2 OLTP versus OLAP	4
1.2.1 OLAP- versus OLTP-Transaktionen	5
1.2.2 Vergleich von OLTP und OLAP	6
1.2.3 Abgrenzung: DBMS-Techniken	7
1.3 Charakteristika und Begriffe	8
1.4 Big Data und Data Warehousing	9
1.5 Aufbau des Buches	11
1.6 Vertiefende Literatur	13
1.7 Übungen	14
2 Architektur	15
2.1 Anforderungen	15
2.1.1 Grobe Übersicht über Data-Warehouse-Systeme	15
2.1.2 Anforderungen an die Architektur	17
2.1.3 Die 12 OLAP-Regeln nach Codd	18
2.1.4 Die FASMI-Anforderungen	21
2.2 Datenfluss in einem Data-Warehouse-System	22
2.2.1 Phasen des Data Warehousing	22
2.2.2 Datenquellen	23
2.2.3 Datenbereinigungsbereich	26
2.2.4 Extraktionskomponenten	27
2.2.5 Transformationskomponenten	27
2.2.6 Ladekomponente	29
2.2.7 Basisdatenbank	29
2.2.8 Befüllen	29
2.2.9 Der Datenwürfel	30

2.2.10	Data Marts	30
2.2.11	Das Data Warehouse	31
2.3	Referenzarchitektur	31
2.3.1	Data-Warehouse-Manager	32
2.3.2	Monitore	33
2.3.3	Repository	35
2.3.4	Metadaten-Manager	35
2.3.5	Diskussion der kompletten Referenzarchitektur	35
2.4	Architektur des Data Warehouse	36
2.4.1	Rolle der Data Marts	36
2.4.2	Abhängige Data Marts: Nabe und Speiche	37
2.4.3	Unabhängige Data Marts	38
2.4.4	Föderierte und virtuelle Datenwürfel	39
2.4.5	Data-Warehouse-Architektur in der Praxis	39
2.4.6	Ein Multi-Schichten-Architekturansatz	40
2.5	Vertiefende Literatur	43
2.6	Übungen	44
3	Modellierung von Data Warehouses	45
3.1	Das multidimensionale Datenmodell	45
3.1.1	Grundbegriffe	45
3.1.2	Dimensionen	47
3.1.3	Fakten und Kennzahlen	48
3.1.4	Schema des multidimensionalen Datenwürfels	50
3.2	Konzeptuelle Modellierung	52
3.2.1	Das ME/R-Modell	53
3.2.2	ADAPT	54
3.3	Relationale Umsetzung	56
3.3.1	Prinzip der relationalen Abbildung	56
3.3.2	Snowflake-Schema	57
3.3.3	Star-Schema	59
3.3.4	Vergleich von Snowflake- und Star-Schema	60
3.3.5	Fact-Constellation-Schema und Galaxie-Schema	61
3.3.6	Alternative Modellierung von Klassifikationshierarchien	64
3.3.7	Vermeidung von Semantikverlusten	65
3.4	Slowly Changing Dimensions	67
3.4.1	Berichtsanforderungen im Data Warehouse	69
3.4.2	Typdefinitionen nach Kimball	72
3.4.3	Realisierungen im Data Warehouse	74
3.5	Vertiefende Literatur	79
3.6	Übungen	80

4	Extraktions-, Transformations- und Ladeprozess	81
4.1	Qualitätsaspekte	82
4.1.1	Der Datenbereinigungsprozess	84
4.1.2	Duplikaterkennung	88
4.1.3	Vergleichsfunktionen	90
4.1.4	Beheben von Datenkonflikten	96
4.2	Der ETL-Prozess	97
4.3	Die Extraktionsphase	99
4.3.1	Extraktionstechniken	100
4.3.2	Methode des Differential Snapshot	101
4.4	Die Transformationsphase	105
4.4.1	Daten- und Schemakonflikte	107
4.4.2	Mappings im Transformationsschritt	109
4.5	Die Ladephase	113
4.5.1	Verwendung des Oracle SQL-Loader	114
4.5.2	Multi-Table-Insert	116
4.6	Alternativer Ansatz: ELT	118
4.7	Vertiefende Literatur	119
4.8	Übungen	120
5	Anfragen an Data-Warehouse-Datenbanken	123
5.1	Einführung und Anforderungen	123
5.2	OLAP-Operationen	125
5.3	SQL-Operationen für das Data Warehouse	128
5.3.1	Relationale Umsetzung multidimensionaler Anfragen	128
5.3.2	CUBE und ROLLUP	134
5.3.3	OLAP-Funktionen in SQL:2003	138
5.4	MDX	145
5.5	Vertiefende Literatur	151
5.6	Übungen	152
6	Speicherung	153
6.1	Speicherung des Datenwürfels: Array vs. Relationen	153
6.1.1	Relationale Implementierung – ROLAP	154
6.1.2	Implementierung als Array – MOLAP	156
6.1.3	Vergleich ROLAP und MOLAP-Speicherung	160
6.1.4	Hybride Speicherung – HOLAP	163
6.1.5	Alternative Speicherungsformen	164
6.2	Partitionierung	167
6.2.1	Partitionierung in relationalen Datenbanken	167
6.2.2	Partitionierung in Data Warehouses	170
6.2.3	Partitionierung von Datenwürfeln	171

6.3	Spaltenorientierte Datenhaltung	176
6.3.1	Basisideen der spaltenorientierten Datenhaltung	176
6.3.2	Operationen und Anfragen in spaltenorientierter Datenhaltung	177
6.3.3	Speichervarianten in spaltenorientierter Datenhaltung	179
6.4	Hauptspeicherdatenbanken	185
6.4.1	Was sind Hauptspeicherdatenbanken?	185
6.4.2	Technologien aktueller Hauptspeicherdatenbanken	187
6.4.3	Komprimierung von Daten	188
6.4.4	Delta-Relationen	190
6.5	Vertiefende Literatur	191
6.6	Übungen	192
7	Indexstrukturen	195
7.1	Klassifikation von Indexstrukturen	195
7.2	B-Bäume und Varianten	198
7.2.1	Der B^+ -Baum	199
7.2.2	Degenerierte B-Bäume	200
7.2.3	Ordnungsabhängigkeit in B-Bäumen	201
7.2.4	B^+ -Baum-Tricks: Oversized Index	202
7.2.5	B^+ -Baum-Tricks: Berechnete Indexe	203
7.3	Bitmap-Indexe	203
7.3.1	Prinzip von Bitmap-Indexen	204
7.3.2	Bitmap-Index: Realisierung	205
7.3.3	Standard-Bitmap-Index	206
7.3.4	Mehrkomponenten-Bitmap-Index	207
7.3.5	Bereichskodierter Bitmap-Index	208
7.3.6	Mehrkomponenten-bereichskodierter Bitmap-Index	208
7.3.7	Intervallkodierte Indexierung	210
7.3.8	Auswahl von Bitmap-Indexstrukturen	211
7.4	Verbundindexe	212
7.4.1	Prinzip des Verbundindex	212
7.4.2	Bitmap-Verbundindex	213
7.5	Mehrdimensionale Indexstrukturen	214
7.5.1	Grid-File	214
7.5.2	Mehrdimensionales Hashen MDH	216
7.5.3	KdB-Baum	217
7.5.4	R-Bäume	218
7.5.5	Varianten von R-Bäumen	220
7.5.6	Der UB-Baum	221
7.6	Indexierung von Hierarchien	226
7.6.1	Kodierung von Hierarchien	226
7.6.2	Mehrdimensionales hierarchisches Clustering	227

7.7	Vertiefende Literatur	228
7.8	Übungen	229
8	Anfrageverarbeitung und materialisierte Sichten	231
8.1	Anfrageplanung	231
8.1.1	Überblick	232
8.1.2	Star-Join-Optimierung	234
8.2	Berechnung des CUBE -Operators	237
8.3	Materialisierte Sichten	241
8.3.1	Anfragebeantwortung mit materialisierten Sichten . . .	242
8.3.2	Auswahl materialisierter Sichten	250
8.3.3	Aktualisierung materialisierter Sichten	252
8.3.4	Materialisierte Sichten in aktuellen DBMS	257
8.4	Vertiefende Literatur	260
8.5	Übungen	261
9	Business-Intelligence-Anwendungen	263
9.1	Business Intelligence	264
9.1.1	Begriffsklärung	264
9.1.2	Knowledge Discovery	268
9.1.3	Datenanalyse	271
9.2	Reporting	272
9.2.1	Balanced Scorecard	275
9.2.2	Navigation im Datenwürfel für Ad-hoc-Reporting	277
9.3	Data Mining im BI-Umfeld	279
9.3.1	Warenkorbanalyse	281
9.3.2	Kunden-Clustering	289
9.3.3	Klassifikationsverfahren	293
9.3.4	Zeitreihenanalyse & Prognose	297
9.3.5	Data Mining Extensions	299
9.4	Vertiefende Literatur	301
9.5	Übungen	302
	Abbildungsverzeichnis	304
	Tabellenverzeichnis	310
	Sachindex	312
	Literaturverzeichnis	319