

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Betriebswirtschaftlicher Ursprung des Data Warehousing .....	2
1.2	Statistischer Ursprung des Data Warehousing .....	5
1.3	Föderativer Ursprung des Data Warehousing .....	7
1.4	Definition eines Data-Warehouse-Systems .....	9
1.5	Ziel und Aufbau des Buches .....	11
<b>2</b>	<b>Aufbau von Data-Warehouse-Systemen</b>	<b>15</b>
2.1	Eigenschaften eines Data-Warehouse-Systems .....	16
2.1.1	Abgrenzung aus der Perspektive der Anwendung .....	16
2.1.2	Abgrenzung aus der Perspektive der Datenhaltung .....	19
2.1.3	Abgrenzung aus der Perspektive des Entwicklungsprozesses ....	20
2.1.4	Zusammenfassung .....	21
2.2	Architektur eines Data-Warehouse-Systems .....	22
2.2.1	Bestandteile eines Data-Warehouse-Systems .....	22
2.2.2	Konfigurationen eines Data-Warehouse-Systems .....	25
2.3	Data-Warehouse-System als Mittel der Integration .....	28
2.3.1	Integration durch Nutzung der Quellsysteme .....	29
2.3.2	Integration durch Datenextraktion und Transformation .....	31
2.3.3	Schritte der Integration in einem Data-Warehouse-System .....	33
2.3.4	Zusammenfassung .....	37

2.4	Data-Warehouse-Systeme als Analyseplattform .....	38
2.4.1	Aufbereitung der Datenbasen .....	38
2.4.2	Nutzung der Data-Warehouse-Datenbasen .....	40
2.5	Metadatenverwaltung in Data-Warehouse-Systemen .....	45
2.5.1	Aufgaben und Rolle der Metadatenverwaltung .....	46
2.5.2	Ebene der Metamodellierung .....	46
2.5.3	Basiskonzepte des CWM-Standards .....	48
2.6	Zusammenfassung .....	51
<b>3</b>	<b>Multidimensionale Datenmodellierung in Data-Warehouse-Systemen</b> .....	<b>53</b>
3.1	Gründe multidimensionaler Modellierung .....	53
3.1.1	Motivation multidimensionaler Modellierung .....	53
3.1.2	Übersicht auswertungsorientierter Modellierungsansätze .....	56
3.1.3	Theorie der funktionalen Abhängigkeit .....	58
3.2	Graphbasierte Modellierungsansätze .....	59
3.2.1	Grundlegende Graphstruktur .....	59
3.2.2	Erweiterungen durch neue Knotentypen .....	61
3.2.3	Knoten zur Modellierung der Abstraktion .....	63
3.3	Grundlagen der multidimensionalen Datenmodellierung .....	64
3.3.1	Dimensionshierarchien .....	65
3.3.2	Fakten, Kennzahlen und Datenwürfel .....	67
3.3.3	Summierbarkeit im multidimensionalen Datenmodell .....	69
3.3.4	Zusammenfassung .....	73
3.4	Zugriff auf multidimensionale Strukturen .....	73
3.4.1	Multidimensionale Operatoren zur Navigation .....	74
3.4.2	Die multidimensionale Anfragesprache MDX .....	76
3.4.3	Die XML-basierte Anfragespezifikation XML Analysis .....	81
3.4.4	Zusammenfassung .....	82

<b>4</b>	<b>Relationale Strukturen und SQL-Operatoren</b>	<b>83</b>
4.1	Das TPC-H/R-Szenario .....	84
4.2	Relationales Star- und Snowflake-Schemamuster .....	85
4.3	Das Muster der Star-Queries .....	90
4.3.1	Auswertungsschritte einer Star-Query .....	91
4.3.2	Gruppierungen .....	92
4.3.3	Aggregationsfunktionen .....	93
4.4	Erweiterte Abbildungstechniken .....	94
4.4.1	Abbildung klassenspezifischer Attribute .....	94
4.4.2	Unbalancierte Hierarchien .....	98
4.4.3	Irreguläre Hierarchien .....	100
4.5	Mehrfachgruppierungen .....	102
4.5.1	Gruppierungsmengen .....	102
4.5.2	Multidimensionale Gruppierung .....	104
4.5.3	Hierarchische multidimensionale Gruppierung .....	106
4.5.4	Partielle multidimensionale Gruppierung .....	108
4.5.5	Zusammenfassung .....	111
4.6	Spezifikation sequenzbasierter Operationen .....	112
4.6.1	Konzept der attributlokalen und tupelbasierten Aggregation .....	112
4.6.2	Attributlokale Partitionierung .....	114
4.6.3	Sequenzorientierte Analysen .....	115
4.6.4	Bildung dynamischer Fenster .....	118
4.6.5	Wechselwirkung mit Mehrfachgruppierungen .....	119
4.6.6	Zusammenfassung .....	121
4.7	Fazit der relationalen Abbildung .....	122

<b>5</b>	<b>Aufbau integrierter Datenbestände</b>	<b>123</b>
5.1	Daten- und Schemaintegration .....	124
5.1.1	Überblick der Daten- und Schemaintegrationsproblematik ....	124
5.1.2	Techniken der Schemaintegration .....	126
5.1.3	Phasen der Datenbereinigung .....	133
5.1.4	Wahrung und Bewertung der Datenqualität .....	134
5.2	Extraktion aus Quellsystemen .....	139
5.2.1	Eigenschaften der Quellsysteme .....	140
5.2.2	Zugriff auf heterogene Datenquellen .....	142
5.2.3	Export und Import .....	143
5.2.4	Externe Tabellen, Tabellenfunktionen und Kapselung .....	147
5.3	Datenübernahmeprozess in einem Data-Warehouse-System .....	151
5.3.1	Explizite Datenübernahme .....	151
5.3.2	Datenübernahme durch Replikation .....	155
5.4	Zusammenfassung .....	158
<b>6</b>	<b>Aufbau konsistenter Datenbestände</b>	<b>161</b>
6.1	Konsistenz in Data-Warehouse-Systemen .....	161
6.1.1	Konsistenzebenen in einem Data-Warehouse-System .....	162
6.1.2	Konsistenzgrade in einem Data-Warehouse-System .....	164
6.1.3	Zusammenfassung .....	165
6.2	Rekonstruktion global konsistenter Zustände .....	166
6.2.1	Aktualisierung einzelner Relationen .....	166
6.2.2	Aktualisierung zusammengesetzter Relationen .....	169
6.3	Gegenseitige Konsistenz .....	173
6.4	Auswahl von Hilfssichten .....	177

6.5	Asynchrone Aktualisierung .....	181
6.5.1	Bestimmung zeitabhängiger Nettoänderungen .....	182
6.5.2	Propagierung durch Vorwärts- und Kompensationsanfragen .....	183
6.5.3	Beispiel zur Bestimmung zeitabhängiger Nettoänderungen ....	184
6.5.4	Algorithmus der asynchronen Aktualisierung .....	186
6.6	Nebenläufige Aktualisierungen .....	187
6.6.1	Externe Kompensation .....	189
6.6.2	Lokale Kompensation des Fehlerterms durch verzögerte Änderungsermittlung .....	191
6.6.3	Lokale Kompensation des Fehlerterms durch iterative Änderungsermittlung .....	193
6.7	Zusammenfassung .....	197
<b>7</b>	<b>Aufbau auswertungsbezogener Datenbestände</b> .....	<b>199</b>
7.1	Unterstützung durch logische Zugriffspfade .....	200
7.1.1	Rahmenmodell .....	200
7.1.2	Partitionierung von Datenbeständen .....	203
7.1.3	Explizite relationale Repräsentation materialisierter Summendaten .....	208
7.1.4	Implizite relationale Abbildung von Summendaten .....	212
7.1.5	Zusammenfassung .....	215
7.2	Voraussetzungen für materialisierte Summendaten .....	216
7.2.1	Theorie der Ableitbarkeit .....	216
7.2.2	Analyse des Problemraums .....	222
7.2.3	Zusammenfassung .....	230
7.3	Auswahl materialisierter Summendaten .....	230
7.3.1	Einflussfaktoren und Kostenmodell .....	231
7.3.2	Statische Auswahlalgorithmen .....	232
7.3.3	Auswahlverfahren unter Berücksichtigung von Aktualisierungskosten .....	237
7.3.4	Dynamische Ansätze zur Auswahl von Summendaten .....	242
7.3.5	Zusammenfassung .....	246

7.4	Pflege materialisierter Summendaten .....	246
7.4.1	Überblick über Aktualisierungsalternativen .....	247
7.4.2	Vollständige Aktualisierung von Summendaten .....	250
7.4.3	Inkrementelle Aktualisierung von Summendaten .....	254
7.4.4	Zusammenfassung .....	260
7.5	Fazit über den Aufbau analyseorientierter Datenbanken .....	260
<b>8</b>	<b>Optimierung von Data-Warehouse-Anfragen</b> .....	<b>263</b>
8.1	Grundlegende Methoden der Anfrageverarbeitung .....	264
8.1.1	Abstrakte Sicht auf die Anfrageverarbeitung .....	264
8.1.2	Interndarstellung einer SQL-Anfrage .....	266
8.1.3	Logische Optimierung .....	269
8.1.4	Physische Optimierung .....	274
8.1.5	Zusammenfassung .....	277
8.2	Nutzung materialisierter Summendaten .....	277
8.3	Realisierungskonzepte des Gruppierungsoperators .....	283
8.3.1	Sortierungsbasierte Realisierung einer einfachen Gruppierung .....	283
8.3.2	Sortierungsbasierte Realisierung von OLAP-Funktionen .....	285
8.3.3	Realisierungskonzepte komplexer Gruppierungs- anweisungen über direkte Ableitbarkeit .....	288
8.3.4	Auflösung komplexer Gruppierungsweisungen durch Gruppierung über Konstanten .....	294
8.3.5	Optimierung komplexer Gruppierungsanweisungen für dünnbesetzte Datenräume .....	298
8.4	Realisierungskonzepte der Scan-Operation .....	301
8.4.1	Einsatz klassischer baumstrukturierter Indexverfahren .....	302
8.4.2	Bitmap-Indexstrukturen .....	305
8.4.3	Multidimensionale Indexstrukturen für Aggregate .....	309
8.4.4	Zusammenfassung .....	316

8.5	Realisierung der Verbundoperation .....	317
8.5.1	Nutzung von Indexstrukturen für den Verbundoperator .....	317
8.5.2	Star-Join-Verbundstrategie .....	320
8.5.3	Realisierungsalternativen der Verbundstrategie bei Partitionen und Parallelität .....	325
8.5.4	Zusammenfassung .....	327
8.6	Fazit und Anwendungsrichtlinien .....	327
8.6.1	Leistungsmessung .....	328
8.6.2	Leistungsoptimierung .....	329
<b>9</b>	<b>Zusammenfassung und Ausblick</b>	<b>333</b>
<b>A</b>	<b>Anhang: Aufbau des TPC-H-Szenarios</b>	<b>335</b>
A.1	Das TPC-H-Schema im Überblick .....	335
A.2	Schemadefinition .....	336
A.3	Spezifikation der Primär-/Fremdschlüsselbeziehungen .....	337
A.4	Definition von Hilfssichten .....	339
A.5	Ladeskript der Datenbank für IBM DB2 V7.2 .....	340
A.6	Aktivieren der Integritätsbedingungen .....	341
	<b>Abbildungsverzeichnis</b>	<b>343</b>
	<b>Tabellenverzeichnis</b>	<b>347</b>
	<b>Literaturverzeichnis</b>	<b>349</b>
	<b>Stichwortverzeichnis</b>	<b>369</b>