

Inhaltsverzeichnis

| | |
|---|------------|
| Inhaltsverzeichnis | vii |
| 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 Aufbau des Buches | 9 |
| 1.5 Vertiefende Literatur | 11 |
| 1.6 Übungen | 11 |
| 2 Architektur | 13 |
| 2.1 Anforderungen | 13 |
| 2.1.1 Grobe Übersicht über Data-Warehouse-Systeme | 13 |
| 2.1.2 Anforderungen an die Architektur | 15 |
| 2.1.3 Die 12 OLAP-Regeln nach Codd | 16 |
| 2.1.4 Die FASMI-Anforderungen | 19 |
| 2.2 Datenfluss in einem Data-Warehouse-System | 20 |
| 2.2.1 Phasen des Data Warehousing | 20 |
| 2.2.2 Datenquellen | 21 |
| 2.2.3 Datenbereinigungsbereich | 24 |
| 2.2.4 Extraktionskomponenten | 25 |
| 2.2.5 Transformationskomponenten | 25 |
| 2.2.6 Ladekomponente | 27 |
| 2.2.7 Basisdatenbank | 27 |
| 2.2.8 Befüllen | 27 |
| 2.2.9 Der Datenwürfel | 28 |
| 2.2.10 Data Marts | 28 |

| | | |
|----------|---|-----------|
| 2.2.11 | Das Data Warehouse | 29 |
| 2.3 | Referenzarchitektur | 29 |
| 2.3.1 | Data-Warehouse-Manager | 30 |
| 2.3.2 | Monitore | 31 |
| 2.3.3 | Repository | 32 |
| 2.3.4 | Metadaten-Manager | 33 |
| 2.3.5 | Diskussion der kompletten Referenzarchitektur | 33 |
| 2.4 | Architektur des Data Warehouse | 33 |
| 2.4.1 | Rolle der Data Marts | 34 |
| 2.4.2 | Abhängige Data Marts: Nabe und Speiche | 35 |
| 2.4.3 | Unabhängige Data Marts | 35 |
| 2.4.4 | Föderierte und virtuelle Datenwürfel | 36 |
| 2.4.5 | Data-Warehouse-Architektur in der Praxis | 37 |
| 2.5 | Vertiefende Literatur | 37 |
| 2.6 | Übungen | 38 |
| 3 | Modellierung von Data Warehouses | 41 |
| 3.1 | Das multidimensionale Datenmodell | 41 |
| 3.1.1 | Grundbegriffe | 41 |
| 3.1.2 | Dimensionen | 43 |
| 3.1.3 | Fakten und Kennzahlen | 44 |
| 3.1.4 | Schema des multidimensionalen Datenwürfels | 46 |
| 3.2 | Konzeptuelle Modellierung | 48 |
| 3.2.1 | Das ME/R-Modell | 49 |
| 3.2.2 | ADAPT | 50 |
| 3.3 | Relationale Umsetzung | 52 |
| 3.3.1 | Prinzip der relationalen Abbildung | 52 |
| 3.3.2 | Snowflake-Schema | 53 |
| 3.3.3 | Star-Schema | 55 |
| 3.3.4 | Vergleich von Snowflake- und Star-Schema | 56 |
| 3.3.5 | Fact-Constellation-Schema und Galaxie-Schema | 57 |
| 3.3.6 | Alternative Modellierung von Klassifikationshierarchien | 60 |
| 3.3.7 | Vermeidung von Semantikverlusten | 61 |
| 3.4 | Vertiefende Literatur | 63 |
| 3.5 | Übungen | 64 |
| 4 | Extraktions-, Transformations- und Ladeprozess | 65 |
| 4.1 | Qualitätsaspekte | 66 |
| 4.1.1 | Der Datenbereinigungsprozess | 68 |
| 4.1.2 | Duplikaterkennung | 72 |
| 4.1.3 | Vergleichsfunktionen | 74 |
| 4.1.4 | Beheben von Datenkonflikten | 79 |
| 4.2 | Der ETL-Prozess | 80 |

| | | |
|----------|--|------------|
| 4.3 | Die Extraktionsphase | 82 |
| 4.3.1 | Extraktionstechniken | 83 |
| 4.3.2 | Methode des Differential Snapshot | 84 |
| 4.4 | Die Transformationsphase | 89 |
| 4.4.1 | Daten- und Schemakonflikte | 90 |
| 4.4.2 | Mappings im Transformationsschritt | 93 |
| 4.5 | Die Ladephase | 96 |
| 4.5.1 | Verwendung des Oracle SQL-Loader | 97 |
| 4.5.2 | Multi-Table-Insert | 100 |
| 4.6 | Alternativer Ansatz: ELT | 101 |
| 4.7 | Vertiefende Literatur | 102 |
| 4.8 | Übungen | 103 |
| 5 | Anfragen an Data-Warehouse-Datenbanken | 107 |
| 5.1 | Einführung und Anforderungen | 107 |
| 5.2 | OLAP-Operationen | 109 |
| 5.3 | SQL-Operationen für das Data Warehouse | 112 |
| 5.3.1 | Relationale Umsetzung multidimensionaler Anfragen | 112 |
| 5.3.2 | CUBE und ROLLUP | 118 |
| 5.3.3 | OLAP-Funktionen in SQL:2003 | 122 |
| 5.4 | MDX | 129 |
| 5.5 | Vertiefende Literatur | 135 |
| 5.6 | Übungen | 136 |
| 6 | Speicherung | 139 |
| 6.1 | Speicherung des Datenwürfels: Array vs. Relationen | 139 |
| 6.1.1 | Relationale Implementierung – ROLAP | 140 |
| 6.1.2 | Implementierung als Array – MOLAP | 142 |
| 6.1.3 | Vergleich ROLAP und MOLAP-Speicherung | 146 |
| 6.1.4 | Hybride Speicherung – HOLAP | 149 |
| 6.1.5 | Alternative Speicherungsformen | 150 |
| 6.2 | Partitionierung | 153 |
| 6.2.1 | Partitionierung in relationalen Datenbanken | 153 |
| 6.2.2 | Partitionierung in Data Warehouses | 156 |
| 6.2.3 | Partitionierung von Datenwürfeln | 157 |
| 6.3 | Spaltenorientierte Datenhaltung | 162 |
| 6.3.1 | Basisideen der spaltenorientierten Datenhaltung | 162 |
| 6.3.2 | Operationen und Anfragen in spaltenorientierter Datenhaltung | 163 |
| 6.3.3 | Speichervarianten in spaltenorientierter Datenhaltung | 165 |
| 6.4 | Hauptspeicherdatenbanken | 171 |
| 6.4.1 | Was sind Hauptspeicherdatenbanken? | 171 |
| 6.4.2 | Technologien aktueller Hauptspeicherdatenbanken | 173 |

| | | |
|----------|--|------------|
| 6.4.3 | Komprimierung von Daten | 174 |
| 6.4.4 | Delta-Relationen | 176 |
| 6.5 | Vertiefende Literatur | 177 |
| 6.6 | Übungen | 178 |
| 7 | Indexstrukturen | 181 |
| 7.1 | Klassifikation von Indexstrukturen | 181 |
| 7.2 | B-Bäume und Varianten | 184 |
| 7.2.1 | Der B^+ -Baum | 185 |
| 7.2.2 | Degenerierte B-Bäume | 186 |
| 7.2.3 | Ordnungsabhängigkeit in B-Bäumen | 187 |
| 7.2.4 | B^+ -Baum-Tricks: Oversized Index | 188 |
| 7.2.5 | B^+ -Baum-Tricks: Berechnete Indexe | 189 |
| 7.3 | Bitmap-Indexe | 189 |
| 7.3.1 | Prinzip von Bitmap-Indexen | 190 |
| 7.3.2 | Bitmap-Index: Realisierung | 191 |
| 7.3.3 | Standard-Bitmap-Index | 192 |
| 7.3.4 | Mehrkomponenten-Bitmap-Index | 193 |
| 7.3.5 | Bereichskodierter Bitmap-Index | 194 |
| 7.3.6 | Mehrkomponenten-bereichskodierter Bitmap-Index | 194 |
| 7.3.7 | Intervallkodierte Indexierung | 196 |
| 7.3.8 | Auswahl von Bitmap-Indexstrukturen | 197 |
| 7.4 | Verbundindexe | 198 |
| 7.4.1 | Prinzip des Verbundindex | 198 |
| 7.4.2 | Bitmap-Verbundindex | 199 |
| 7.5 | Mehrdimensionale Indexstrukturen | 200 |
| 7.5.1 | Grid-File | 200 |
| 7.5.2 | Mehrdimensionales Hashen MDH | 202 |
| 7.5.3 | KdB-Baum | 203 |
| 7.5.4 | R-Bäume | 204 |
| 7.5.5 | Varianten von R-Bäumen | 206 |
| 7.5.6 | Der UB-Baum | 207 |
| 7.6 | Indexierung von Hierarchien | 212 |
| 7.6.1 | Kodierung von Hierarchien | 212 |
| 7.6.2 | Mehrdimensionales hierarchisches Clustering | 213 |
| 7.7 | Vertiefende Literatur | 214 |
| 7.8 | Übungen | 215 |
| 8 | Anfrageverarbeitung und materialisierte Sichten | 217 |
| 8.1 | Anfrageplanung | 217 |
| 8.1.1 | Überblick | 218 |
| 8.1.2 | Star-Join-Optimierung | 220 |
| 8.2 | Berechnung des CUBE-Operators | 223 |

| | | |
|----------|--|------------|
| 8.3 | Materialisierte Sichten | 227 |
| 8.3.1 | Anfragebeantwortung mit materialisierten Sichten . . . | 228 |
| 8.3.2 | Auswahl materialisierter Sichten | 236 |
| 8.3.3 | Aktualisierung materialisierter Sichten | 238 |
| 8.3.4 | Materialisierte Sichten in aktuellen DBMS | 243 |
| 8.4 | Vertiefende Literatur | 246 |
| 8.5 | Übungen | 247 |
| 9 | Business-Intelligence-Anwendungen | 249 |
| 9.1 | Business Intelligence | 250 |
| 9.1.1 | Begriffsklärung | 250 |
| 9.1.2 | Knowledge Discovery | 255 |
| 9.1.3 | Datenanalyse | 257 |
| 9.2 | Reporting | 259 |
| 9.2.1 | Balanced Scorecard | 260 |
| 9.2.2 | Navigation im Datenwürfel für Ad-hoc-Reporting | 263 |
| 9.3 | Data Mining im BI-Umfeld | 265 |
| 9.3.1 | Warenkorbanalyse | 266 |
| 9.3.2 | Kunden-Clustering | 275 |
| 9.3.3 | Klassifikationsverfahren | 278 |
| 9.3.4 | Zeitreihenanalyse & Prognose | 283 |
| 9.3.5 | Data Mining Extensions | 285 |
| 9.4 | Vertiefende Literatur | 287 |
| 9.5 | Übungen | 287 |
| | Abbildungsverzeichnis | 290 |
| | Tabellenverzeichnis | 295 |
| | Sachindex | 296 |
| | Literaturverzeichnis | 303 |