

Andreas Bauer · Holger Günzel (Hrsg.)

Data-Warehouse-Systeme

Architektur · Entwicklung · Anwendung

2., überarbeitete und aktualisierte Auflage



dpunkt.verlag

Inhaltsverzeichnis

Teil I	Architektur	1
1	Abgrenzung und Einordnung	5
1.1	Begriffliche Einordnung	5
1.1.1	Definitionen	7
1.1.2	Abgrenzung von transaktionalen Systemen	8
1.2	Historie des Themenbereichs	11
1.3	Anwendungsbereiche	13
1.3.1	Informationsorientierte Anwendungen	14
1.3.2	Analyseorientierte Anwendungen	17
1.3.3	Planungsorientierte Anwendungen	21
1.3.4	Kampagnenorientierte Anwendungen	23
1.4	Einführung in das Beispiel <i>Star*Kauf</i>	25
1.5	Überblick über das Buch	28
2	Referenzarchitektur	31
2.1	Aspekte einer Referenzarchitektur	31
2.1.1	Referenzmodell für die Architektur von Data-Warehouse-Systemen	32
2.1.2	Beschreibung der Referenzarchitektur	34
2.2	Data-Warehouse-Manager	37
2.3	Datenquelle	38
2.3.1	Bestimmung der Datenquellen	39
2.3.2	Datenqualität	43
2.3.3	Klassifikation der Quelldaten	45
2.4	Monitor	47
2.5	Arbeitsbereich	48
2.6	Extraktionskomponente	49
2.7	Transformationskomponente	50

2.8	Ladekomponente	51
2.9	Basisdatenbank	51
2.9.1	Charakterisierung, Aufgaben und Abgrenzung	52
2.9.2	Aktualisierungsalternativen der Basisdatenbank	55
2.9.3	Qualität der Daten in der Basisdatenbank	55
2.10	Data Warehouse	57
2.10.1	Unterstützung des Ladeprozesses	58
2.10.2	Unterstützung des Analyseprozesses	58
2.10.3	Data Marts – Verteilung des Data Warehouse	59
2.11	Analyse	63
2.11.1	Darstellungsformen	64
2.11.2	Funktionalität	65
2.11.3	Realisierung	66
2.11.4	Plattformen	68
2.12	Repository	68
2.13	Metadatenmanager	70
2.14	Zusammenfassung	71
3	Phasen des Data Warehousing	75
3.1	Monitoring	75
3.1.1	Realisierungen des Monitoring	76
3.1.2	Monitoring-Techniken	77
3.2	Extraktionsphase	81
3.3	Transformationsphase	83
3.3.1	Datenintegration	83
3.3.2	Bereinigung	89
3.4	Ladephase	94
3.5	Analysephase	97
3.5.1	Data Access	97
3.5.2	Online Analytical Processing (OLAP)	97
3.5.3	Data Mining	109
3.6	Zusammenfassung	117
4	Physische Architektur	119
4.1	Speicherarchitekturen für das Data Warehouse und die Basisdatenbank	119
4.1.1	Architektur eines Datenbankverwaltungssystems	120
4.1.2	Speichermodelle für Daten	120

4.2	Schichtenarchitekturen	122
4.2.1	Einschichtenarchitektur	124
4.2.2	Zweischichtenarchitektur	124
4.2.3	Dreischichtenarchitektur	126
4.2.4	N-Schichtenarchitektur	126
4.2.5	Web-basierte Architektur	127
4.3	Middleware	132
4.3.1	Normen und Spezifikationen	132
4.3.2	Middleware-Systeme	132
4.3.3	Common Object Request Broker Architecture (CORBA) ..	134
4.3.4	Web-Services	136
4.3.5	Distributed Component Object Model (DCOM)	138
4.3.6	.NET	139
4.3.7	Java 2 Enterprise Edition	141
4.4	Schnittstellen	143
4.4.1	Klassifikation von Schnittstellen	143
4.4.2	Techniken und Standards	144
4.4.3	Datenaustauschformate	146
4.5	Sicherheit	148
4.5.1	Kommunikationssicherheit	149
4.5.2	Benutzeridentifikation und Authentifizierung	151
4.5.3	Auditing	152
4.5.4	Zugriffskontrolle	153
4.6	Zusammenfassung	155
Teil II	Entwicklung	157
5	Das multidimensionale Datenmodell	161
5.1	Konzeptuelle Modellierung	162
5.1.1	Verschiedene Vorgehensweisen zur Definition einer Methodik	164
5.1.2	Vorstellung verschiedener Designnotationen	165
5.2	Logische Modellierung	174
5.2.1	Notwendigkeit der Formalisierung des multidimen- sionalen Modells	175
5.2.2	Struktur des multidimensionalen Datenmodells	176
5.2.3	Fehlende Werte in Würfelzellen (Nullwerte)	180
5.2.4	Operatoren des multidimensionalen Modells	181
5.2.5	Weitere Ansätze zur Formalisierung	185
5.2.6	Grenzen und Erweiterungen des multidimensionalen Datenmodells	187

5.3	Unterstützung von Veränderungen	188
5.3.1	Zeitaspekte	188
5.3.2	Aspekte der Klassifikationsveränderungen	190
5.3.3	Aspekte der Schemaänderung	193
5.4	Zusammenfassung	200
6	Umsetzung des multidimensionalen Datenmodells	201
6.1	Relationale Speicherung	202
6.1.1	Abbildungsmöglichkeiten auf Relationen	202
6.1.2	Relationale Umsetzung multidimensionaler Anfragen	214
6.1.3	Relationale Versionierungs- und Evolutionsaspekte	220
6.2	Multidimensionale Speicherung	225
6.2.1	Datenstrukturen	226
6.2.2	Speicherung multidimensionaler Daten	235
6.2.3	Dateneingabe	239
6.2.4	Grenzen der multidimensionalen Datenhaltung	241
6.2.5	Hybride Speicherung; Hybrides OLAP (HOLAP)	242
6.3	Realisierung der Zugriffskontrolle	243
6.3.1	Zugriffskontrollanforderungen	244
6.3.2	Realisierungskonzepte	248
6.4	Zusammenfassung	252
7	Optimierung	253
7.1	Anfragen im multidimensionalen Modell	254
7.2	Indexstrukturen	255
7.2.1	Klassifikation der Indexstrukturen	256
7.2.2	Eindimensionale Baumindexstrukturen	258
7.2.3	Mehrdimensionale Baumindexstrukturen	263
7.2.4	Bitmap-Indizes	269
7.2.5	Vergleich der Indizierungstechniken	275
7.3	Partitionierung	277
7.3.1	Horizontale Partitionierung	277
7.3.2	Vertikale Partitionierung	280
7.3.3	Mini-Dimensionen als Spezialfall vertikaler Partitionierung	281
7.3.4	Partitionierungssteuerung	282
7.4	Relationale Optimierung von Star-Joins	283

7.5	Einsatz materialisierter Sichten	286
7.5.1	Verwendung materialisierter Sichten	288
7.5.2	Bestimmung des Auswertekontextes für Aggregatanfragen	292
7.5.3	Statische Auswahl materialisierter Sichten	297
7.5.4	Dynamische Auswahl materialisierter Sichten	305
7.5.5	Aktualisierung materialisierter Sichten	308
7.6	Optimierung eines multidimensionalen Datenbanksystems	318
7.6.1	Partitionierung	319
7.6.2	Speicherung der Zellen	323
7.6.3	Datenblockindizierung	324
7.7	Zusammenfassung	325
8	Metadaten	327
8.1	Die Rolle von Metadaten beim Data Warehousing	327
8.2	Metadatenmanagement	330
8.2.1	Anforderungen an Data-Warehouse-Repositoryn	330
8.2.2	Repository- und Metadatenaustauschstandards	333
8.3	Data-Warehouse-Metadatenschemata	335
8.3.1	Eine Klassifikation für Metadaten	336
8.3.2	Standards und Referenzmodelle	338
8.4	Entwurf eines Schemas zur Verwaltung von Data-Warehouse-Metadaten	342
8.4.1	Funktionale Aspekte	343
8.4.2	Personen, Organisation und Aufgaben	345
8.4.3	Business-Metadaten	345
8.4.4	Abstraktionsstufen	346
8.5	Zusammenfassung	347
Teil III	Anwendung	349
9	Vorgehensweise beim Aufbau eines Data-Warehouse-Systems	353
9.1	Data-Warehouse-Strategie	353
9.1.1	IT-Strategie	354
9.1.2	Data-Warehouse-Strategie	356
9.1.3	Rolle des Data Warehousing innerhalb der IT-Strategie ..	356
9.2	Ableitung der Data-Warehouse-Architektur	357
9.2.1	Data-Warehouse-Rahmenwerk als gesamtheitliche Vorgabe	358
9.2.2	Umgang mit mehreren Data-Warehouse-Systemen	361
9.2.3	Architekturüberlegungen in der Praxis	364
9.2.4	Umgebungen im Hinblick auf Entwicklung, Test, Produktion und Wartung	366

9.3	Data-Warehouse-Vorgehensweise	368
9.3.1	Phasenkonzept	369
9.3.2	Vorgehensweisen bei der Einführung	372
9.3.3	Machbarkeitsbetrachtung zum Data Warehousing	376
9.3.4	Analysephase	377
9.3.5	Designphase	381
9.3.6	Implementierungsphase	382
9.4	Zusammenfassung	385
10	Das Data-Warehouse-Projekt	387
10.1	Data-Warehouse-Projektorganisation	387
10.1.1	Projekttrollen und Projektsteuerung	388
10.1.2	Projektteam und Rollenverständnis	389
10.1.3	Kommunikation	392
10.1.4	Konfliktmanagement	393
10.1.5	Qualitätssicherung	394
10.1.6	Dokumentation	395
10.2	Softwareauswahl	396
10.2.1	Nutzen und Notwendigkeit der Produktauswahl	396
10.2.2	Klassifikation der Produkte anhand der Referenzarchitektur	397
10.2.3	Vorgehensweise zur Produktauswahl	399
10.2.4	Allgemeine Kriterien für die Produktauswahl	406
10.2.5	Kriterien für Datenbeschaffungswerkzeuge	407
10.2.6	Kriterien für OLAP-Produkte	412
10.3	Hardwareauswahl	417
10.3.1	Auswahlbestimmende Faktoren	418
10.3.2	Datenspeicherung	419
10.3.3	Archivspeichermedien	420
10.3.4	Multiprozessorsysteme	421
10.3.5	Fehlertoleranz als Planungsziel	423
10.3.6	Flaschenhälse und Fallstricke	424
10.3.7	Backup-Strategien und Notfallpläne	424
10.4	Erfolgsfaktoren beim Aufbau eines Data-Warehouse-Systems	426
10.4.1	Institutionelle Aufgaben des Projektmanagements: Projektorganisation	426
10.4.2	Funktionale Aufgaben des Projektmanagements: Projektabwicklung	428
10.4.3	Empfehlungen für ein Data-Warehouse-Projekt	430
10.5	Wirtschaftlichkeitsbetrachtungen	431
10.5.1	Kostenbetrachtung	432
10.5.2	Nutzenbetrachtung	433
10.6	Zusammenfassung	437

11	Betrieb eines Data-Warehouse-Systems	439
11.1	Administration	439
11.1.1	Anforderungen und resultierende Aufgaben	440
11.1.2	Organisationsformen für den Betrieb	447
11.2	Iterativer Datenbeschaffungsprozess	449
11.3	Performanz-Tuning von Data-Warehouse-Systemen	455
11.3.1	Der Performanz-Tuning-Prozess	455
11.3.2	Maßnahmen aus Sicht des Informationsmanagements	456
11.3.3	Maßnahmen aus Sicht des Datenbankdesigns	457
11.3.4	Maßnahmen aus Sicht der Applikationsumgebung	460
11.3.5	Maßnahmen aus Sicht der Datenbankzugriffe	462
11.3.6	Maßnahmen aus Sicht der Datenbankkonfiguration	463
11.3.7	Maßnahmen aus Sicht des Betriebssystems	465
11.3.8	Maßnahmen aus Sicht des Netzwerks	466
11.3.9	Maßnahmen aus Sicht des Hardwaresystems	467
11.4	Analyseprozess	468
11.4.1	Schere zwischen Systemleistung und Anwendererwartungen	468
11.4.2	Anwenderbetreuung	471
11.4.3	Tipps und Tricks im Umgang mit Anwendern	473
11.5	Rolle des Repositoriums	473
11.6	Sicherungsmanagement	475
11.6.1	Backup und Recovery	475
11.6.2	Entsorgung von Daten	477
11.6.3	Datenbank- und Systemverfügbarkeit	479
11.6.4	Phasen eines Recovery-Plans	480
11.7	Zusammenfassung	481
12	Praxisbeispiele	483
12.1	Data Warehousing in der Gesundheitsberichterstattung	484
12.1.1	Das Epidemiologische Krebsregister Niedersachsen (EKN)	484
12.1.2	Unterstützung umweltbezogener epidemiologischer Forschung durch modernes Data Warehousing	485
12.1.3	Fazit	491
12.2	Praxisbeispiel aus dem Verlagswesen	491
12.2.1	Vorstellung des Axel Springer Verlags	492
12.2.2	Das Data-Warehouse-Projekt im Anzeigenbereich	492
12.2.3	Fazit	497

12.3	Panelorientierte Marktforschung	498
12.3.1	Das Unternehmen GfK AG und die GfK Marketing Services GmbH	499
12.3.2	Data Warehousing in der panelorientierten Marktforschung	499
12.3.3	Fazit	504
12.4	Grundlagenforschung: Geowissenschaften	505
12.4.1	Das ICDP-Projekt und sein Träger	505
12.4.2	Data Warehousing in den Geowissenschaften	506
12.4.3	Fazit	511
12.5	Chemie: Distribution von Produktdaten	512
12.5.1	Das Unternehmen Merck KGaA	512
12.5.2	Data Warehousing für Produktdaten	513
12.5.3	Fazit	517
12.6	Zusammenfassung	518
Anhang		519
A	Abkürzungen	521
B	Glossar	525
C	Autorenverzeichnis	531
D	Autorenuordnung	541
E	Literatur und Web-Referenzen	545
	Stichwortverzeichnis	577