

Andreas Bauer • Holger Günzel (Hrsg.)

Data-Warehouse- Systeme

Architektur • Entwicklung • Anwendung

3., überarbeitete und aktualisierte Auflage

HOCHSCHULE
LIECHTENSTEIN
Bibliothek



dpunkt.verlag

Inhaltsverzeichnis

• rv, , ;

&i

r ; >"'.'!.

Teil I	Architektur	1
1	Abgrenzung und Einordnung	S
1.1	Begriffliche Einordnung	6
1.1.1	Definitionen	7
1.1.2	Abgrenzung von transaktionalen Systemen	9
1.2	Historie des Themenbereichs	11
1.3	Einordnung und Abgrenzung von Business Intelligence	13
1.4	Anwendungsbereiche	14
1.4.1	Informationsorientierte Anwendungen	16
1.4.2	Analyseorientierte Anwendungen	18
1.4.3	Planungsorientierte Anwendungen	22
1.4.4	Kampagnenorientierte Anwendungen	24
1.5	Einführung in das Beispiel <i>Star*Kauf</i>	26
1.6	Überblick über das Buch	28
2	Referenzarchitektur	33
2.1	Aspekte einer Referenzarchitektur	33
2.1.1	Referenzmodell für die Architektur von Data-Warehouse-Systemen	34
2.1.2	Beschreibung der Referenzarchitektur	36
2.2	Data-Warehouse-Manager	39
2.3	Datenquelle	40
2.3.1	Bestimmung der Datenquellen	41
2.3.2	Datenqualität	45
2.3.3	Klassifikation der Quelldaten	47
2.4	Monitor	49
2.5	Arbeitsbereich	50

4	Physische Architektur	125
4.1	Speicherarchitekturen für das Data Warehouse und die Basisdatenbank	125
4.1.1	Architektur eines Datenbankverwaltungssystems	126
4.1.2	Speichermodelle für Daten	126
4.2	Schichtenarchitekturen	128
4.2.1	Einschichtenarchitektur	130
4.2.2	Zweischichtenarchitektur	130
4.2.3	Dreischichtenarchitektur	132
4.2.4	N-Schichtenarchitektur	132
4.2.5	Webbasierte Architektur	133
4.3	Middleware	138
4.3.1	Normen und Spezifikationen	138
4.3.2	Middleware-Systeme	138
4.3.3	Common Object Request Broker Architecture (CORBA)	140
4.3.4	Webservices	142
4.3.5	Distributed Component Object Model (DCOM)	144
4.3.6	.NET	145
4.3.7	Java 2 Enterprise Edition	147
4.4	Schnittstellen	149
4.4.1	Klassifikation von Schnittstellen	149
4.4.2	Techniken und Standards	150
4.4.3	Datenaustauschformate	152
4.4.4	Anfragesprache MDX	153
4.5	Sicherheit	159
4.5.1	Netzwerksicherheit	159
4.5.2	Benutzeridentifikation und Authentifizierung	162
4.5.3	Auditing	163
4.5.4	Autorisierung und Zugriffskontrolle	164
4.6	Zusammenfassung	168

Teil II Entwicklung

5	Das multidimensionale Datenmodell	173
5.1	Konzeptuelle Modellierung	174
5.1.1	Verschiedene Vorgehensweisen zur Definition einer Methodik	176
5.1.2	Vorstellung verschiedener Designnotationen	177

5.2	Logische Modellierung	186
> l	5.2.1 Notwendigkeit der Formalisierung des multidimensionalen Modells	187
. _	5.2.2 Struktur des multidimensionalen Datenmodells	188
i	5.2.3 Fehlende Werte in Würfelzellen (Nullwerte)	192
iii	5.2.4 Operatoren des multidimensionalen Modells	193
"f- 5	5.2.5 Weitere Ansätze zur Formalisierung	197
/i: i	5.2.6 Grenzen und Erweiterungen des multidimensionalen Datenmodells	199
5.3	Unterstützung von Veränderungen	200
7Jj	5.3.1 Zeitaspekte	200
	5.3.2 Aspekte der Klassifikationsveränderungen	202
	5.3.3 Aspekte der Schemaänderung	205
5.4	Zusammenfassung	212
4	Umsetzung des multidimensionalen Datenmodells	213
6.1	Relationale Speicherung	214
-'*-	6.1.1 Abbildungsmöglichkeiten auf Relationen	214
^ * *	6.1.2 Relationale Umsetzung multidimensionaler Anfragen	226
	6.1.3 Relationale Versionierungs- und Evolutionsaspekte	232
6.2	Multidimensionale Speicherung	237
.	6.2.1 Datenstrukturen	238
;' "	6.2.2 Speicherung multidimensionaler Daten	247
	6.2.3 Dateneingabe	251
	6.2.4 Grenzen der multidimensionalen Datenhaltung	253
	6.2.5 Hybride Speicherung: Hybrides OLAP (HOLAP)	254
6.3	Realisierung der Zugriffskontrolle	255
	6.3.1 Zugriffskontrolleanforderungen	255
	6.3.2 Relationale Realisierung	258
	6.3.3 Multidimensionale Realisierung	260
	6.3.4 Inferenzen und Trackerangriffe	262
	6.3.5 Realisierungskonzepte	263
6.4	Zusammenfassung	267
7	Optimierung	269
7.1	Anfragen im multidimensionalen Modell	270
7.2	Indexstrukturen	271
	7.2.1 Klassifikation der Indexstrukturen	272
	7.2.2 Eindimensionale Baumindexstrukturen	274
• *	7.2.3 Mehrdimensionale Baumindexstrukturen	279
	7.2.4 Bitmap-Indizes	285
	7.2.5 Vergleich der Indizierungstechniken	290

Inhaltsverzeichnis

7.3	Partitionierung	292
7.3.1	Horizontale Partitionierung	293
7.3.2	Vertikale Partitionierung	295
7.3.3	Mini-Dimensionen als Spezialfall vertikaler Partitionierung	296
7.3.4	Partitionierungssteuerung	297
7.3.5	Multidimensional Clustering	298
7.4	Relationale Optimierung von Star-Joins.	301
7.5	Einsatz materialisierter Sichten	305
7.5.1	Verwendung materialisierter Sichten.	306
7.5.2	Bestimmung des Auswertekontextes für Aggregatanfragen.	311
7.5.3	Statische Auswahl materialisierter Sichten.	315
7.5.4	Dynamische Auswahl materialisierter Sichten	323
7.5.5	Aktualisierung materialisierter Sichten	326
7.6	Optimierung eines multidimensionalen Datenbanksystems.	336
7.6.1	Partitionierung	337
7.6.2	Speicherung der Zellen	340
7.6.3	Datenblockindizierung	342
7.7	Zusammenfassung	343
8	Metadaten	345
8.1	Metadaten und Metamodelle beim Data Warehousing	345
8.2	Metadatenmanagement	348
8.3	Metadatenmanagementsystem - Metadaten-Warehouse	350
8.3.1	Anforderungen an ein Metadatenmanagementsystem	351
8.3.2	Architektur.	352
8.3.3	Repositorium- und Metadatenaustauschstandards.	355
8.4	Data-Warehouse-Metadaten schemata	358
8.4.1	Eine Klassifikation für Metadaten	359
8.4.2	Standards und Referenzmodelle	361
8.5	Entwurf eines Schemas zur Verwaltung von Data-Warehouse-Metadaten.	365
8.5.1	Funktionale Aspekte.	366
8.5.2	Personen, Organisation und Aufgaben	368
8.5.3	Business-Metadaten.	368
8.5.4	Abstraktionsstufen.	369
8.6	Zusammenfassung	370

Teil III Anwendung

9	Vorgehensweise beim Aufbau eines Data-Warehouse-Systems	377
9.1	Data-Warehouse-Strategie	378
9.1.1	IT-Strategie	378
9.1.2	Data-Warehouse-Strategie	380
9.1.3	Rolle des Data Warehouse innerhalb der IT-Strategie	380
9.2	Reifegradmodell	381
9.3	Ableitung der Data-Warehouse-Architektur	387
9.3.1	Data-Warehouse-Rahmenwerk als gesamtheitliche Vorgabe	387
9.3.2	Umgang mit mehreren Data-Warehouse-Systemen	391
9.3.3	Architekturüberlegungen in der Praxis	394
9.3.4	Umgebungen im Hinblick auf Entwicklung, Test, Produktion und Wartung	396
9.4	Data-Warehouse-Vorgehensweise	398
9.4.1	Grundsätzliche Überlegungen zum Projektvorgehen	399
9.4.2	Phasenkonzept	400
9.4.3	Vorgehensweisen bei der Einführung	404
9.4.4	Machbarkeitsbetrachtung zum Data Warehousing	408
9.4.5	Analysephase	409
9.4.6	Designphase	413
9.4.7	Implementierungsphase . . . *-<- . ** . v	418
9.5	Zusammenfassung	421
10	Das Data-Warehouse-Projekt	423
10.1	Data-Warehouse-Projektorganisation	423
10.1.1	Projektrollen und Projektsteuerung	424
10.1.2	Projektteam und Rollenverständnis	425
10.1.3	Kommunikation	428
10.1.4	Konfliktmanagement	429
10.1.5	Qualitätssicherung	430
10.1.6	Dokumentation	431
10.2	Softwareaus wähl	432
10.2.1	Nutzen und Notwendigkeit der Produktauswahl	432
10.2.2	Klassifikation der Produkte anhand der Referenzarchitektur	433
**	10.2.3 Vorgehensweise zur Produktauswahl	436
ii.	10.2.4 Allgemeine Kriterien für die Produktauswahl	443
	10.2.5 Kriterien für DatenbeschaffungsWerkzeuge	444
	10.2.6 Kriterien für OLAP-Produkte	448

10.3	Hardwareauswahl	454
10.3.1	Auswahlbestimmende Faktoren.	454
10.3.2	Datenspeicherung	455
10.3.3	Archivspeichermedien.	457
10.3.4	Multiprozessorsysteme.	458
10.3.5	Fehlertoleranz als Planungsziel.	461
10.3.6	Flaschenhalse und Fallstricke.	461
10.3.7	Backup-Strategien und Notfallplane.	462
10.4	Erfolgsfaktoren beim Aufbau eines Data-Warehouse-Systems	463
10.4.1	Institutionelle Aufgaben des Projektmanagements: Projektorganisation	464
10.4.2	Funktionale Aufgaben des Projektmanagements: Projektabwicklung	465
10.4.3	Empfehlungen fur ein Data-Warehouse-Projekt	467
10.5	WirtschaftlichkeitsBetrachtungen	468
10.5.1	Kostenbetrachtung	469
10.5.2	Nutzenbetrachtung	470
10.6	Zusammenfassung	475
11	Betrieb und Weiterentwicklung eines Data-Warehouse-Systems	477
11.1	Administration	477
11.1.1	Anforderungen und resultierende Aufgaben	478
11.1.2	Organisationsformen fur Entwicklung und Betrieb.	485
11.1.3	Rolle des Repositoriums.	488
11.2	Datenbeschaffungsprozess.	490
11.3	Performanz-Tuning von Data-Warehouse-Systemen	496
11.3.1	Der Performanz-Tuning-Prozess.	496
11.3.2	Manahmen aus Sicht des Informationsmanagements	497
11.3.3	Manahmen aus Sicht des Datenbankdesigns.	498
11.3.4	Manahmen aus Sicht der Applikationsumgebung	501
11.3.5	Manahmen aus Sicht der Datenbankzugriffe.	502
11.3.6	Manahmen aus Sicht der Datenbankkonfiguration.	504
11.3.7	Manahmen aus Sicht des Betriebssystems.	506
11.3.8	Manahmen aus Sicht des Netzwerks.	507
11.3.9	Manahmen aus Sicht des Hardwaresystems.	508
11.3.10	Multicore-Architekturen.	508
11.4	Analyseprozess.	512
11.4.1	Schere zwischen Systemleistung und Anwendererwartungen.	512
11.4.2	Anwenderbetreuung	515

11.5	Sicherungsmanagement	517
f ••	11.5.1 Backup und Recovery	517
'i >	11.5.2 Entsorgung von Daten	519
'-T ••	11.5.3 Datenbank- und Systemverfügbarkeit	521
	11.5.4 Phasen eines Recovery-Plans	522
11.6	Zusammenfassung	523
12	Praxisbeispiele	525
12.1	Öffentliche Verwaltung	526
	12.1.1 Die Bundesagentur für Arbeit	526
t .it.	12.1.2 Data Warehousing in der öffentlichen Arbeitsverwaltung	527
	12.1.3 Fazit	534
12.2	Data Warehousing in der Versicherung	535
	12.2.1 Risikomanagement auf Basis eines Data-Warehouse- Systems in einem Versicherungskonzern	535
	12.2.2 Fazit	540
12.3	Panelorientierte Marktforschung	541
	12.3.1 Die GfK-Gruppe und die GfK Marketing Services GmbH & Co. KG	542
≥ 'r.	12.3.2 Data Warehousing in der panelorientierten Marktforschung	542
	12.3.3 Fazit	547
12.4	Chemie und Pharma	548
rV:J-	12.4.1 Das Unternehmen Merck KGaA	548
•'•«j,	12.4.2 Konzernweites Monitoring von ERP-Prozessen in einem Verbund von Data-Warehouse-Systemen	549
-	12.4.3 Fazit	556
12.5	Zusammenfassung	557
 nnang		
A	Abkürzungen	561
	Glossar	565
	Autorenverzeichnis • •>	571
	Autorenuordnung	581
E	Literatur und Webreferenzen	»
	Stichwortverzeichnis	619