

Datenbanksysteme

Eine Einführung

von

Professor Alfons Kemper, Ph.D.

Dr. Andre Eickler

5., aktualisierte und erweiterte Auflage

Oldenbourg Verlag München Wien

Inhaltsverzeichnis

Vorwort	15
1 Einleitung und Übersicht	17
1.1 Motivation für den Einsatz eines DBMS.	17
1.2 Datenabstraktion	19
1.3 Datenunabhängigkeit	20
1.4 Datenmodelle	21
1.5 Datenbankschema und Ausprägung	22
1.6 Einordnung der Datenmodelle	22
1.6.1 Modelle des konzeptuellen Entwurfs	22
1.6.2 Logische (Implementations-)Datenmodelle	23
1.7 Architekturübersicht eines DBMS	26
1.8 Übungen	28
1.9 Literatur	28
2 Datenbankentwurf	29
2.1 Abstraktionsebenen des Datenbankentwurfs	29
2.2 Allgemeine Entwurfsmethodik	30
2.3 Die Datenbankentwurfsschritte	31
2.4 Die Anforderungsanalyse	31
2.4.1 Informationsstrukturanforderungen	33
2.4.2 Datenverarbeitungsanforderungen	35
2.5 Grundlagen des Entity-Relationship-Modells	35
2.6 Schlüssel	37
2.7 Charakterisierung von Beziehungstypen	37
2.7.1 Funktionalitäten der Beziehungen	37
2.7.2 Funktionalitätsangaben bei n-stelligen Beziehungen	39
2.7.3 Die (<i>min</i> , <i>max</i>)-Notation	42
2.8 Existenzabhängige Entitytypen	46
2.9 Generalisierung	47
2.10 Aggregation	48
2.11 Kombination von Generalisierung und Aggregation	50
2.12 Konsolidierung, Sichtenintegration	51
2.13 Konzeptuelle Modellierung mit UML	57
2.13.1 UML-Klassen	57
2.13.2 Assoziationen zwischen Klassen	58
2.13.3 Aggregation in UML	59
2.13.4 Anwendungsbeispiel: Begrenzungsflächendarstellung von Polyedern in UML	60
2.13.5 Generalisierung in UML-Notation	61
2.13.6 Die Modellierung der Universität in UML	61
2.13.7 Verhaltensmodellierung in UML	62
2.13.8 Anwendungsfall-Modellierung (use cases)	62

2.13.9	Interaktionsdiagramme	64
2.13.10	Interaktionsdiagramm zur Prüfungsdurchführung	64
2.14	Übungen	65
2.15	Literatur	67
3	Das relationale Modell	69
3.1	Definition des relationalen Modells	69
3.1.1	Mathematischer Formalismus	69
3.1.2	Schema-Definition	70
3.2	Umsetzung eines konzeptuellen Schemas in ein relationales Schema	71
3.2.1	Relationale Darstellung von Entitytypen	71
3.2.2	Relationale Darstellung von Beziehungen	71
3.3	Verfeinerung des relationalen Schemas	76
3.3.1	1 ¹ -Beziehungen	76
3.3.2	1:1-Beziehungen	78
3.3.3	Relationale Modellierung der Generalisierung	79
3.3.4	Beispielausprägung der Universitäts-Datenbank	80
3.3.5	Relationale Modellierung schwacher Entitytypen	80
3.4	Die relationale Algebra	82
3.4.1	Selektion	82
3.4.2	Projektion	83
3.4.3	Vereinigung	83
3.4.4	Mengendifferenz	84
3.4.5	Kartesisches Produkt (Kreuzprodukt)	84
3.4.6	Umbenennung von Relationen und Attributen	85
3.4.7	Definition der relationalen Algebra	86
3.4.8	Der relationale Verbund (Join)	87
3.4.9	Mengendurchschnitt	90
3.4.10	Die relationale Division	92
3.4.11	Operatorbaum-Darstellung	93
3.5	Der Relationenkalkül	93
3.5.1	Beispielanfrage im relationalen Tupelkalkül	94
3.5.2	Quantifizierung von Tupelvariablen	95
3.5.3	Formale Definition des Tupelkalküls	96
3.5.4	Sichere Ausdrücke des Tupelkalküls	97
3.5.5	Der relationale Domänenkalkül	97
3.5.6	Beispielanfragen im Domänenkalkül	98
3.5.7	Sichere Ausdrücke des Domänenkalküls	99
3.6	Ausdruckskraft der Anfragesprachen	100
3.7	Übungen	100
3.8	Literatur	103
4	Relationale Anfragesprachen	105
4.1	Geschichte	105
4.2	Datentypen	106
4.3	Schemadefinition	106
4.4	Schemaveränderung	107
4.5	Elementare Datenmanipulation: Einfügen von Tupeln	107
4.6	Einfache SQL-Anfragen	108
4.7	Anfragen über mehrere Relationen	109

Inhaltsverzeichnis

4.8	Aggregatfunktionen und Gruppierung	112
4.9	Geschachtelte Anfragen	114
4.10	Quantifizierte Anfragen in SQL	118
4.11	Nullwerte	120
4.12	Spezielle Sprachkonstrukte	121
4.13	Joins in SQL-92	123
4.14	Rekursion	123
4.15	Veränderungen am Datenbestand	128
4.16	Sichten	129
4.17	Sichten zur Modellierung von Generalisierungen	130
4.18	Charakterisierung update-fähiger Sichten	132
4.19	Einbettung von SQL in Wirtssprachen	133
4.20	Anfragen in Anwendungsprogrammen	135
4.21	JDBC: Java Database Connectivity	137
4.21.1	Verbindungsaufbau zu einer Datenbank	138
4.21.2	Resultset-Programmbeispiel	140
4.21.3	Vorübersetzung von SQL-Ausdrücken	140
4.22	SQLJ: Eine Einbettung von SQL in Java	142
4.23	Query by Example	143
4.24	Übungen	147
4.25	Literatur	149
Datenintegrität		151
5.1	Referentielle Integrität	152
5.2	Gewährleistung referentieller Integrität	152
5.3	Referentielle Integrität in SQL	153
5.4	Überprüfung statischer Integritätsbedingungen	154
5.5	Das Universalitätsschema mit Integritätsbedingungen	156
5.6	Komplexere Integritätsbedingungen	158
5.7	Trigger	159
5.8	Übungen	161
5.9	Literatur	162
Relationale Entwurfstheorie		163
6.1	Funktionale Abhängigkeiten	163
6.1.1	Konventionen zur Notation	164
6.1.2	Einhaltung einer funktionalen Abhängigkeit	164
6.2	Schlüssel	165
6.3	Bestimmung funktionaler Abhängigkeiten	166
6.3.1	Kanonische Überdeckung	169
6.4	„Schlechte" Relationenschemata	170
6.4.1	Die Updateanomalien	170
6.4.2	Einfügeanomalien	171
6.4.3	Löschanomalien	171
6.5	Zerlegung (Dekomposition) von Relationen	171
6.5.1	Verlustlosigkeit	172
6.5.2	Kriterien für die Verlustlosigkeit einer Zerlegung	174
6.5.3	Abhängigkeitsbewahrung	175
6.6	Erste Normalform	177
6.7	Zweite Normalform	178

6.8	Dritte Normalform	180
6.9	Boyce-Codd Normalform	182
6.10	Mehrwertige Abhängigkeiten	185
6.11	Vierte Normalform	187
6.12	Zusammenfassung	189
6.13	Übungen	189
6.14	Literatur	193
Physische Datenorganisation		195
7.1	Speichermedien	195
7.2	Speicherarrays: RAID	196
7.3	Der Datenbankpuffer	200
7.4	Abbildung von Relationen auf den Sekundärspeicher	202
7.5	Indexstrukturen	203
7.6	ISAM	205
7.7	B-Bäume	207
7.8	B ⁺ -Bäume	211
7.9	Präfix-B ⁺ -Bäume	212
7.10	Hashing	213
7.11	Erweiterbares Hashing	215
7.12	Mehrdimensionale Indexstrukturen	217
7.13	Ballung logisch verwandter Datensätze	221
7.14	Unterstützung eines Anwendungsverhaltens	223
7.15	Physische Datenorganisation in SQL	225
7.16	Übungen	225
7.17	Literatur	227
Anfragebearbeitung		229
8.1	Logische Optimierung	230
8.1.1	Äquivalenzen in der relationalen Algebra	232
8.1.2	Anwendung der Transformationsregeln	234
8.2	Physische Optimierung	238
8.2.1	Implementierung der Selektion	240
8.2.2	Implementierung von binären Zuordnungsoperatoren	240
8.2.3	Gruppierung und Duplikateliminiierung	247
8.2.4	Projektion und Vereinigung	248
8.2.5	Zwischenspeicherung	248
8.2.6	Übersetzung der logischen Algebra	250
8.3	Kostenmodelle	253
8.3.1	Selektivitäten	254
8.3.2	Kostenabschätzung für die Selektion	256
8.3.3	Kostenabschätzung für den Join	257
8.3.4	Kostenabschätzung für die Sortierung	257
8.4	„Tuning“ von Datenbankabfragen	258
8.5	Übungen	259
8.6	Literatur	262

9	Transaktionsverwaltung	265
9.1	Begriffsbildung	265
9.2	Anforderungen an die Transaktionsverwaltung	266
9.3	Operationen auf Transaktions-Ebene	266
9.4	Abschluss einer Transaktion	267
9.5	Eigenschaften von Transaktionen	269
9.6	Transaktionsverwaltung in SQL	270
9.7	Zustandsübergänge einer Transaktion	271
9.8	Literatur	272
10	Fehlerbehandlung	273
10.1	Fehlerklassifikation	273
10.1.1	Lokaler Fehler einer Transaktion	273
10.1.2	Fehler mit Hauptspeicherverlust	274
10.1.3	Fehler mit Hintergrundspeicherverlust	275
10.2	Die Speicherhierarchie	275
10.2.1	Ersetzung von Puffer-Seiten	275
10.2.2	Einbringen von Änderungen einer Transaktion	276
10.2.3	Einbringstrategie	277
10.2.4	Hier zugrunde gelegte Systemkonfiguration	278
10.3	Protokollierung von Änderungsoperationen	278
10.3.1	Struktur der Log-Einträge	279
10.3.2	Beispiel einer Log-Datei	279
10.3.3	Logische oder physische Protokollierung	279
10.3.4	Schreiben der Log-Information	280
10.3.5	Das WAL-Prinzip	282
10.4	Wiederanlauf nach einem Fehler	282
10.4.1	Analyse des Logs	283
10.4.2	Redo-Phase	284
10.4.3	Undo-Phase	284
10.5	Fehlertoleranz des Wiederanlaufs	284
10.6	Lokales Zurücksetzen einer Transaktion	286
10.7	Partielles Zurücksetzen einer Transaktion	287
10.8	Sicherungspunkte	288
10.8.1	Transaktionskonsistente Sicherungspunkte	288
10.8.2	Aktionskonsistente Sicherungspunkte	289
10.8.3	Unschärfe (fuzzy) Sicherungspunkte	291
10.9	Recovery nach einem Verlust der materialisierten Datenbasis	292
10.10	Übungen	293
10.11	Literatur	294
11	Mehrbenutzersynchronisation	295
11.1	Fehler bei unkontrolliertem Mehrbenutzerbetrieb	296
11.1.1	Verlorengegangene Änderungen (<i>lost Update</i>)	296
11.1.2	Abhängigkeit von nicht freigegebenen Änderungen	296
11.1.3	Phantomproblem	297
11.2	Serialisierbarkeit	297
11.2.1	Beispiele serialisierbarer Ausführungen (Historien)	298
11.2.2	Nicht serialisierbare Historie	298
11.3	Theorie der Serialisierbarkeit	301

11.3.1	Definition einer Transaktion	301
11.3.2	Historie (Schedule)	302
11.3.3	Äquivalenz zweier Historien	303
11.3.4	Serialisierbare Historien	304
11.3.5	Kriterien für Serialisierbarkeit	304
11.4	Eigenschaften von Historien bezüglich der Recovery	306
11.4.1	Rücksetzbare Historien	306
11.4.2	Historien ohne kaskadierendes Rücksetzen	306
11.4.3	Strikte Historien	307
11.4.4	Beziehungen zwischen den Klassen von Historien	307
11.5	Der Datenbank-Scheduler	308
11.6	Sperrbasierte Synchronisation	309
11.6.1	Zwei Sperrmodi	309
11.6.2	Zwei-Phasen-Sperrprotokoll	310
11.6.3	Kaskadierendes Rücksetzen (Schneeballeffekt)	312
11.7	Verklemmungen (Deadlocks)	312
11.7.1	Erkennung von Verklemmungen	313
11.7.2	Preclaiming zur Vermeidung von Verklemmungen	314
11.7.3	Verklemmungsvermeidung durch Zeitstempel	315
11.8	Hierarchische Sperrgranulate	316
11.9	Einfüge- und Löschooperationen, Phantome	320
11.10	Zeitstempel-basierende Synchronisation	321
11.11	Optimistische Synchronisation	323
11.12	Synchronisation von Indexstrukturen	324
11.13	Mehrbenutzersynchronisation in SQL-92	327
11.14	Übungen	329
11.15	Literatur	332
12	Sicherheitsaspekte	333
12.1	Discretionary Access Control	335
12.2	Zugriffskontrolle in SQL	335
12.2.1	Identifikation und Authentisierung	336
12.2.2	Autorisierung und Zugriffskontrolle	336
12.2.3	Sichten	337
12.2.4	Auditing	338
12.3	Verfeinerung des Autorisierungsmodells	338
12.3.1	Implizite Autorisierung von Subjekten	339
12.3.2	Implizite Autorisierung von Operationen	340
12.3.3	Implizite Autorisierung von Objekten	341
12.3.4	Implizite Autorisierung entlang einer Typhierarchie	341
12.4	Mandatory Access Control	343
12.5	Multilevel-Datenbanken	343
12.6	Kryptographie	347
12.6.1	Der Data Encryption Standard	347
12.6.2	Public-Key Kryptographie	349
12.7	Zusammenfassung	350
12.8	Übungen	351
12.9	Literatur	352

13 Objektorientierte Datenbanken	353
13.1 Bestandsaufnahme relationaler Datenbanksysteme	353
13.2 Vorteile der objektorientierten Datenmodellierung	357
13.3 Der ODMG-Standard	358
13.4 Eigenschaften von Objekten	359
13.4.1 Objektidentität	360
13.4.2 Typ eines Objekts	361
13.4.3 Wert eines Objekts.	361
13.5 Definition von Objekttypen	362
13.5.1 Attribute.	362
13.5.2 Beziehungen	362
13.5.3 Typeigenschaften: Extensionen und Schlüssel.	369
13.6 Modellierung des Verhaltens: Operationen.	369
13.7 Vererbung und Subtypisierung	372
13.7.1 Terminologie.	372
13.7.2 Einfache und Mehrfach Vererbung.	373
13.8 Beispiel einer Typhierarchie.	374
13.9 Verfeinerung (Spezialisierung) und spätes Binden von Operationen	377
13.10 Mehrfachvererbung	380
13.11 Die Anfragesprache OQL	381
13.11.1 Einfache Anfragen	381
13.11.2 Geschachtelte Anfragen und Partitionierung	382
13.11.3 Pfadausdrücke.	383
13.11.4 Erzeugung von Objekten	384
13.11.5 Operationsaufruf.	384
13.12 C++-Einbettung	384
13.12.1 Objektidentität	386
13.12.2 Objekterzeugung und Ballung.	387
13.12.3 Einbettung von Anfragen.	387
13.13 Übungen	388
13.14 Literatur.	389
14 Erweiterbare und objekt-relationale Datenbanken	391
14.1 Übersicht über die objekt-relationalen Konzepte.	391
14.2 Large Objects (LOBs).	392
14.3 Distinct Types: Einfache benutzerdefinierte Datentypen.	394
14.4 Table Functions.	398
14.4.1 Nutzung einer <i>Table Function</i> in Anfragen.	399
14.4.2 Implementierung einer <i>Table Function</i>	399
14.5 Benutzerdefinierte strukturierte Objekttypen	401
14.6 Geschachtelte Objekt-Relationen	405
14.7 Vererbung von SQL-Objekttypen.	409
14.8 Komplexe Attribut-Typen.	412
14.9 Übungen	413
14.10 Literatur.	414

15	Deduktive Datenbanken	415
15.1	Terminologie	415
15.2	Datalog	415
15.3	Eigenschaften von Datalog-Programmen	419
15.3.1	Rekursivität	419
15.3.2	Sicherheit von Datalog-Regeln	419
15.4	Auswertung von nicht-rekursiven Datalog-Programmen	420
15.4.1	Auswertung eines Beispielprogramms	420
15.4.2	Auswertungs-Algorithmus	423
15.5	Auswertung rekursiver Regeln	425
15.6	Inkrementelle (semi-naive) Auswertung rekursiver Regeln	427
15.7	Bottom-Up oder Top-Down Auswertung	431
15.8	Negation im Regelrumpf	433
15.8.1	Stratifizierte Datalog-Programme	433
15.8.2	Auswertung von Regeln mit Negation	434
15.8.3	Ein etwas komplexeres Beispiel	435
15.9	Ausdruckskraft von Datalog	435
15.10	Übungen	437
15.11	Literatur	440
16	Verteilte Datenbanken	443
16.1	Terminologie und Abgrenzung	443
16.2	Entwurf verteilter Datenbanken	445
16.3	Horizontale und vertikale Fragmentierung	447
16.3.1	Horizontale Fragmentierung	448
16.3.2	Abgeleitete horizontale Fragmentierung	450
16.3.3	Vertikale Fragmentierung	451
16.3.4	Kombinierte Fragmentierung	453
16.3.5	Allokation für unser Beispiel	454
16.4	Transparenz in verteilten Datenbanken	455
16.4.1	Fragmentierungstransparenz	455
16.4.2	Allokationstransparenz	456
16.4.3	Lokale Schema-Transparenz	456
16.5	Anfrageübersetzung und -Optimierung in VDBMS	457
16.5.1	Anfragebearbeitung bei horizontaler Fragmentierung	457
16.5.2	Anfragebearbeitung bei vertikaler Fragmentierung	459
16.6	Join-Auswertung in VDBMS	461
16.6.1	Join-Auswertung ohne Filterung	461
16.6.2	Join-Auswertung mit Filterung	462
16.7	Transaktionskontrolle in VDBMS	465
16.8	Mehrbenutzersynchronisation in VDBMS	470
16.8.1	Serialisierbarkeit	470
16.8.2	Das Zwei-Phasen-Sperrprotokoll in VDBMS	470
16.9	Deadlocks in VDBMS	471
16.9.1	Erkennung von Deadlocks	471
16.9.2	Deadlock-Vermeidung	474
16.10	Synchronisation bei replizierten Daten	475
16.11	Übungen	478
16.12	Literatur	481

17 Betriebliche Anwendungen: OLTP, Data Warehouse, Data Mining	483
17.1 SAP R/3: Ein betriebswirtschaftliches Datenbankanwendungssystem	483
17.1.1 Architektur von SAP R/3	483
17.1.2 Datenmodell und Schema von SAP R/3	484
17.1.3 ABAP/4	485
17.1.4 Transaktionen in SAP R/3	488
17.2 Data Warehouse, Decision-Support, OLAP	489
17.2.1 Datenbankentwurf für das Data Warehouse	490
17.2.2 Anfragen im Sternschema: Star Join	493
17.2.3 Roll-Up/Drill-Down-Anfragen	494
17.2.4 Flexible Auswertungsmethoden	496
17.2.5 Materialisierung von Aggregaten	496
17.2.6 Der cube-Operator	498
17.2.7 Wiederverwendung materialisierter Aggregate	498
17.2.8 Bitmap-Indexe für OLAP-Anfragen	501
17.2.9 Auswertungsalgorithmen für komplexe OLAP-Anfragen	502
17.2.10 Data Warehouse-Architekturen	504
17.3 Data Mining	505
17.3.1 Klassifikation von Objekten	506
17.3.2 Assoziationsregeln	507
17.3.3 Der Ä Priori-Algorithmus	508
17.3.4 Bestimmung der Assoziationsregeln	510
17.3.5 Cluster-Bestimmung	511
17.4 Übungen	512
17.5 Literatur	513
18 Internet-Datenbankanbindungen	515
18.1 HTML- und HTTP-Grundlagen	515
18.1.1 HTML: Die Hypertext-Sprache des World Wide Web	515
18.1.2 Adressierung von Web-Dokumenten	516
18.1.3 Client/Server-Architektur des World Wide Web	518
18.1.4 HTTP: Das HyperText Transfer Protokoll	518
18.1.5 HTTPS	519
18.2 Web-Datenbank-Anbindung via Servlets	520
18.2.1 Beispiel-Servlet	520
18.3 Java Server Pages / Active Server Pages	526
18.3.1 JSP/HTML-Seite mit Java-Code	527
18.3.2 HTML-Seite mit Java-Bean-Aufruf	529
18.3.3 Die Java-Bean Komponente <i>VorlesungenBean</i>	530
18.3.4 Sokrates' Homepage	532
18.4 Datenbankanbindung via Java-Applets	532
18.5 Übungen	533
18.6 Literatur	534
19 XML-Datenmodellierung und Web-Services	535
19.1 XML-Datenmodellierung	535
19.1.1 Schema oder kein Schema	536
19.1.2 Rekursive Schemata	538
19.1.3 Universitätsinformation in XML-Format	538
19.1.4 XML-Namensräume	540

19.1.5	XML Schema: Eine Schemadefinitionssprache	542
19.1.6	Verweise (Referenzen) in XML-Daten	544
19.2	XQuery: Eine XML-Anfragesprache	545
19.2.1	Pfadausdrücke	545
19.2.2	Verkürzte XPath-Syntax	550
19.2.3	Beispiel-Pfadausdrücke in verkürzter Syntax	551
19.2.4	Anfragesyntax von XQuery	552
19.2.5	Geschachtelte Anfragen	554
19.2.6	Joins in XQuery	554
19.2.7	Join-Prädikat im Pfadausdruck	555
19.2.8	Das let-Konstrukt	556
19.2.9	Dereferenzierung in FLWOR-Ausdrücken	557
19.2.10	Rekursive Anfragen	558
19.3	Web-Services	560
19.3.1	Erstellen und Nutzen eines Web-Services im Überblick	562
19.3.2	Das Auffinden von Diensten	564
19.3.3	Ein Beispiel-Web-Service	567
19.3.4	Definition der Web-Service Schnittstellen	567
19.3.5	Nachrichtenformat für die Interaktion mit Web-Services	570
19.3.6	Implementierung des Web-Services	572
19.3.7	Aufruf des Web-Services	573
19.4	Übungen	575
19.5	Literatur	577
20	Leistungsbewertung	579
20.1	Überblick über Datenbanksystem-Benchmarks	579
20.2	Der TPC-C Benchmark	579
20.3	Die TPC-H und TPC-R (früher TPC-D) Benchmarks	582
20.4	Der OO7 Benchmark für oo-Datenbanken	588
20.5	Der TPC-W Benchmark	589
20.6	Übungen	592
20.7	Literatur	592
	Literaturverzeichnis	593
	Index	626