

Datenmodelle, Datenbanksprachen und Datenbank- managementsysteme

von
Prof. Dr. Gottfried Vossen

5. überarbeitete und erweiterte Auflage

Oldenbourg Verlag München Wien

Inhaltsverzeichnis

Auszug ...	xiv
... aus dem Vorwort zur ersten Auflage	xiv
... aus dem Vorwort zur zweiten Auflage	xv
... aus dem Vorwort zur dritten Auflage	xvi
Vorwort zur fünften Auflage	xix
I Einführung	1
1 Motivationen, Historisches, Abgrenzungen	3
1.1 Daten-intensive Anwendungen und deren Anforderungen	3
1.2 Datenbanken und Informationssysteme.	6
1.3 Entwicklung von Datenbanksystemen und Datenmodellen.	7
1.3.1 Filesysteme.	8
1.3.2 Generische Sicht eines Datenbanksystems.	10
1.3.3 Netzwerk- und hierarchische Datenbanken.	12
1.3.4 Relationale Datenbanken.	15
1.3.5 Objektrelationale Datenbanken.	18
1.3.6 XML-Datenbanken.	21
1.4 Bibliographische Hinweise und Ergänzungen.	22
1.5 Übungen	23
2 Aufbau von Datenbanken und Datenbanksystemen	25
2.1 Datenunabhängigkeit	25
2.2 Die ANSI/SPARC-Schemaebenen.	26
2.3 Sprachebenen und -klassen.	29
2.4 Ein Schichtenmodell für Datenbanksysteme.	32
2.5 Bibliographische Hinweise, Ergänzungen, Querbezüge.	39
2.6 Übungen	41

II	Klassische Konzepte:	
	ER- und relationales Datenmodell. SQL	43
3	Datenbank-Entwurf mit dem Entity-Relationship-Modell	45
3.1	Der Datenbank-Lebenszyklus	46
3.2	Aspekte der Qualitätssicherung	49
3.2.1	Qualitätskriterien für den Datenbankentwurf	49
3.2.2	Die Grundzüge ordnungsmäßiger Modellierung	51
3.3	Phasen des Entwurfsprozesses	52
3.3.1	Anforderungsanalyse und -Spezifikation	54
3.3.2	Konzeptioneller Entwurf	55
3.3.3	Logischer Entwurf	57
3.3.4	Physischer Entwurf	57
3.3.5	Weitere Entwurfsschritte	58
3.4	Allgemeine Abstraktionskonzepte	59
3.5	Das Entity-Relationship-Modell	60
3.5.1	Entities und Attribute	61
3.5.2	Relationships	69
3.5.3	IS-A-Beziehungen	75
3.5.4	Zusammenfassung	80
3.6	Konzeptioneller Entwurf mit dem ER-Modell	81
3.7	Bibliographische Hinweise und Ergänzungen	88
3.8	Übungen	90
4	Das relationale Datenmodell	95
4.1	Relationen	95
4.2	Datenabhängigkeiten. Relationenschemata	98
4.3	Transformation eines ER-Diagramms in das Relationenmodell	104
4.4	Datenbank-Definition mit SQL	109
4.4.1	Datenbankschema-Definition	110
4.4.2	Tabellen-Definition	111
4.4.3	Integritätsbedingungen	112
4.4.4	Weitere Definitionsaspekte	114
4.5	Fallstudie Medienhandel	116
4.6	Bibliographische Hinweise und Ergänzungen	124
4.7	Übungen	126
5	Grundlagen von SQL	127
5.1	SQL von 1980 bis 2003	127
5.2	Datenbank-Updates	128
5.3	Datenbank-Anfragen	130
5.3.1	Einführung	130
5.3.2	Zur Syntax von SELECT	131
5.3.3	Exkurs in die Relationenalgebra	134
5.3.4	Semantik von SELECT am Beispiel von RA-Ausdrücken	135
5.3.5	Verbund-Operatoren in SQL	137
5.4	Beispiele für SQL-Anfragen	140
5.4.1	Selektionen und Projektionen	141

5.4.2	Verbundanfragen	142
5.4.3	Aggregatfunktionen und Unteranfragen	144
5.4.4	Abschließende Bemerkungen	146
5.5	Rekursive Anfragen	148
5.6	Bibliographische Hinweise und Ergänzungen	149
5.7	Übungen	149
5.8	Anhang: Beispieltabellen der Fallstudie Medienhandel	153
6	Integrität in relationalen Datenbanken	161
6.1	Arten von Integritätsbedingungen	161
6.2	Funktionale Abhängigkeiten	165
6.2.1	Definition. Implikation	166
6.2.2	Ableitung. Das Membership-Problem	169
6.2.3	Anwendungen des Membership-Algorithmus. Schlüssel	175
6.2.4	(Abhängigkeits-) Basen	178
6.2.5	Allgemeine Implikation	181
6.3	Inklusionsabhängigkeiten	183
6.4	Allgemeine Assertions	185
6.5	Aktive Integritätskontrolle über Trigger	188
6.6	Bibliographische Hinweise, Ergänzungen, Querbezüge	193
6.7	Übungen	198
7	Relationale Sprachen: Relationenalgebra, Kalküle, Datalog	201
7.1	Die Relationenalgebra	201
7.1.1	Relationale Operationen und ihre Rechenregeln	202
7.1.2	Ausdrücke der Relationenalgebra	208
7.1.3	Eigenschaften der Relationenalgebra	210
7.1.4	Algebraische Optimierung	212
7.2	Relationenkalküle	214
7.2.1	Der Relationen-Tupelkalkül	215
7.2.2	Der Relationen-Domainkalkül	220
7.3	View-Updates	221
7.4	Grundlagen von Datalog	226
7.4.1	Motivationen	226
7.4.2	Logik als Anfragesprache	228
7.4.3	Intensionale und extensionale Datenbanken	229
7.4.4	Rekursion	230
7.4.5	Integritätsbedingungen	232
7.4.6	Datalog-Programme und deren Interpretation	232
7.4.7	Ausdruckskraft von Datalog	235
7.5	Konzepte der Theorie von Anfragesprachen	236
7.5.1	Konjunktive Anfragen	236
7.5.2	Vollständigkeit von Anfragesprachen	237
7.5.3	Ausdruckskraft und Komplexität von Anfragesprachen	239
7.6	Bibliographische Hinweise und Ergänzungen	243
7.7	Übungen	246

8	Qualitätsorientierter Datenbankentwurf	251
8.1	Universalrelationen und Update-Anomalien	251
8.2	Zweite, dritte und Boyce-Codd-Normalform	253
8.3	Dekomposition und Synthese	259
8.4	Mehrwertige und Verbund-Abhängigkeiten	269
8.5	Vierte und fünfte Normalform	273
8.6	Bibliographische Hinweise und Ergänzungen	275
8.7	Übungen	277
III Moderne Konzepte:		
Objekt-Basierung und XML-Dokumente		279
9	Objekt-relationale Datenbanken	281
9.1	Evolution von Datenmodellen	282
9.1.1	Grenzen relationaler Datenmodelle	282
9.1.2	Höhere Datenmodelle	283
9.1.3	Daten-, Funktions- und Prozessentwurf	285
9.2	Objekt-Orientierung	288
9.2.1	Das Paradigma	288
9.2.2	Objekt-Orientierung und Datenbanken	292
9.2.3	Unterscheidung von Typen und Klassen	295
9.2.4	Objekt-Orientierung im relationalen Datenmodell	299
9.3	Datenbankmodellierung mit der UML	301
9.3.1	Klassen	303
9.3.2	Assoziationen	303
9.3.3	Aggregationen und Kompositionen	305
9.3.4	Generalisierung und Spezialisierung	305
9.3.5	Weitere Modellierungsmöglichkeiten	306
9.3.6	Fallstudie Medienhandel	307
9.4	Objekt-relationale Definitionskonzepte in SQL	310
9.4.1	Basisdatentypen und Typkonstruktoren	310
9.4.2	Benutzerdefinierte Typen und Typhierarchien	312
9.4.3	Tabellen	315
9.5	Fallstudie Medienhandel	317
9.6	SQL-Anfragen an objekt-relationale Datenbanken	321
9.7	Bibliographische Hinweise und Ergänzungen	326
9.8	Übungen	327
9.9	Anhang: Beispieltabellen der Fallstudie Medienhandel	330
10	Einführung in XML	337
10.1	Elemente, Tags und Attribute	338
10.2	Validierung von XML-Dokumenten: DTDs und XML Schema	345
10.3	Normalisierungsaspekte	349
10.4	Verarbeitung von XML-Dokumenten	351
10.4.1	XPath	352
10.4.2	XSLT	355
10.5	Bibliographische Hinweise und Ergänzungen	363

10.6	Übungen		364
10.7	Anhang: XML-Spezifikationen für die Fallstudie Medienhandel		365
10.7.1	Document Type Definition	<	365
10.7.2	XML Schema Definition		367
11	SQL und Programmiersprachen	->	375
11.1	Web-Anbindung von Datenbanken		376
11.1.1	Grundlegendes	!	376
11.1.2	Scripting		378
11.1.3	MySQL-Anbindung per PHP		380
11.2	Eingebettetes SQL		386
11.2.1	Arten der Einbettung		386
11.2.2	JDBC und SQLJ		388
11.3	Bibliographische Hinweise und Ergänzungen		394
11.4	Übungen		394
12	XQuery und SQL/XML		397
12.1	XQuery		397
12.1.1	Ausdrücke		398
12.1.2	FLWOR-Ausdrücke		400
12.1.3	Funktionen		407
12.1.4	Update-Operationen		408
12.2	SQL/XML		409
12.2.1	XML-Datentypen		411
12.2.2	Publikations-Funktionen		413
12.2.3	XQuery-Funktionen		416
12.3	SQL/XML-Unterstützung in IBM DB2 V9		418
12.3.1	Speicherung von XML-Dateien		420
12.3.2	Anfragen mit SQL/XML		422
12.3.3	Anfragen mit XQuery		424
12.3.4	Generierung von XML- oder HTML-Strukturen		424
12.4	Bibliographische Hinweise und Ergänzungen		425
12.5	Übungen		426
IV	Datenintegration und ihre Anwendungen		435
13	Architekturkonzepte für Datenbanksysteme		437
13.1	Das Client/Server-Konzept		438
13.1.1	Anfrageserver		440
13.1.2	Objekt- und Seitenserver		442
13.2	Verteilte Datenbanksysteme		444
13.2.1	Vor- und Nachteile		444
13.2.2	Klassifikationen		446
13.2.3	Homogene Verteilung		448
13.2.4	Heterogene Verteilung und Föderation		451
13.3	Parallele Datenbanksysteme		454
13.4	Bibliographische Hinweise und Ergänzungen		458

13.5	Übungen	459
14	Datenintegration	461
14.1	Einführung. Architekturkonzepte	461
14.2	Übersetzer, Integratoren und Koordinatoren	463
14.3	Integration mittels logischer Sichten	467
14.4	Verteilte Objekte und Web Services	473
14.5	Bibliographische Hinweise und Ergänzungen	476
14.6	Übungen	478
15	Data Warehouses (Datenlager) und OLAP	479
15.1	Grundlagen	479
15.1.1	Anwendungen und Architektur von Datenlagern	481
15.1.2	Der ETL-Prozess	484
15.1.3	Daten-Bereinigung zur Erzielung von Datenqualität	485
15.2	Mehrdimensionale Datenmodelle	486
15.2.1	Fakten, Maße und Dimensionen	486
15.2.2	Sternschemata und Varianten	488
15.2.3	Effiziente Speicherurig von Faktentabellen mittels Bitmaps	490
15.3	Qualitätsorientierter Schemaentwurf für Datenlager	493
15.3.1	Der Entwurfsprozess	494
15.3.2	Qualitätskriterien	497
15.3.3	Summierbarkeit und Normalformen	497
15.3.4	Selbstwartbarkeit	499
15.4	Grundlagen des OLAP	502
15.4.1	Datenwürfel und OLAP-Operatoren	503
15.4.2	Von Vereinigungs- zu Würfel-Anfragen	506
15.4.3	SQL Roll-Up	510
15.4.4	SQL Cube	513
15.4.5	SQL Grouping Sets	516
15.5	Bibliographische Hinweise und Ergänzungen	517
15.6	Übungen	518
16	Grundlagen des Data Mining	523
16.1	Einführung	524
16.2	Klassifikation	527
16.3	Cluster-Bildung	529
16.4	Assoziationsanalyse	532
16.4.1	Häufige Mengen und Assoziationsregeln	532
16.4.2	Der Apriori-Algorithmus	535
16.4.3	Verbesserungspotenziale beim Apriori-Algorithmus	537
16.4.4	Alternativen zum Apriori-Algorithmus	539
16.4.5	Interessantheitsmaße	543
16.5	Sequenzanalyse	544
16.6	Web-Mining	548
16.6.1	Web-Struktur-Mining	548
16.6.2	Web-Inhalts-Mining	551
16.6.3	Web-Log-Mining	551

16.7 Bibliographische Hinweise und Ergänzungen	552
16.8 Übungen	554

¹ V Datenbanksystemtechnik 555

17 Interne Datenbank- und Speicherorganisation	557
17.1 Plattenspeicher	557
17.1.1 Blöcke und Blockzugriffe	559
17.1.2 RAID-Architekturen	560
17.2 Pufferverwaltung	563
17.3 Files	564
17.3.1 Sequentielle Files	565
17.3.2 Effizienter Filezugriff durch Indexierung	567
17.4 Spezielle Indexstrukturen	568
17.4.1 ISAM	568
17.4.2 Baumstrukturen für eindimensionale Punktdaten	569
17.4.3 Baumstrukturen für mehrdimensionale Punktdaten	575
17.4.4 Hash-Organisationsformen	581
17.5 Beispiel: Speicherorganisation bei DB2	584
17.6 Speicherung und Indexierung von XML-Dokumenten	587
17.6.1 Speicherungsoptionen	587
17.6.2 Indexierung	588
A1.7 Bibliographische Hinweise und Ergänzungen	593
'17.8 Übungen	595
18 Verarbeitung und Optimierung von Anfragen	597
18.1 Anfrageverarbeitung im Überblick	598
18.2 Anfrage-Optimierung	600
18.2.1 Ziele und Ansätze	600
18.2.2 High-Level-Optimierung	601
18.2.3 Low-Level-Optimierung	606
18.3 Implementierungstechniken für den Verbund	607
18.3.1 Nested-Loop-Join	608
18.3.2 Sort-Merge-Join	608
18.3.3 Hash-Join	609
18.4 Spezielle QEP-Parameter	612
18.5 Anfrage-Prozessoren	615
18.6 Bibliographische Hinweise und Ergänzungen	617
18.7 Übungen	618
19 Physischer Datenbankentwurf und Tuning	621
19.1 Grundlagen des physischen Datenbankentwurfs	621
19.2 Fragmentierung und Replikation bei Datenverteilung	624
19.3 Datenbank-Tuning	626
19.3.1 Grundlegende Prinzipien	627
19.3.2 Tuning von Indexen	629
19.3.3 Tuning in relationalen Systemen	629

19.4	Datenbank-Benchmarks	630
19.5	Bibliographische Hinweise und Erganzungen	633
19.6	ubungen	634
20	Transaktionen und Serialisierbarkeit ^N	637
20.1	Das Transaktionskonzept	638
20.2	Beispiele fur Synchronisationsprobleme	641
20.3	Das ACID-Prinzip	643
20.4	Das Read-Write-Modell fur Transaktionen	644
20.5	Serialisierbarkeit von Schedules	647
20.5.1	Schedules	647
20.5.2	View-Serialisierbarkeit	651
20.5.3	Konflikt-Serialisierbarkeit	652
20.5.4	Commit-Serialisierbarkeit	657
20.6	Fehlersicherheit von Schedules	659
20.6.1	Motivation	659
20.6.2	Rucksetzbarkeit	660
20.6.3	Vermeidung kaskadierender Aborts	661
20.6.4	Striktheit	662
20.7	Korrektheit von Schedules	663
20.8	Bibliographische Hinweise und Erganzungen	663
20.9	ubungen	664
21	Concurrency Control	667
21.1	uberlegungen zum Scheduler-Entwurf	667
21.2	Sperrende Scheduler	670
21.2.1	Einfuhrung	670
21.2.2	Das 2-Phasen-Sperrprotokoll	672
21.2.3	Varianten des 2-Phasen-Sperrprotokolls	677
21.2.4	Das MGL-Protokoll	679
21.2.5	Das TL-Protokoll	682
21.2.6	Tuning in Gegenwart sperrender Scheduler	684
21.3	Nicht sperrendes Timestamp-Ordering	689
21.4	Concurrency Control auf hoherer Abstraktionsebene	690
21.5	Bibliographische Hinweise und Erganzungen	694
21.6	ubungen	695
22	Grundlagen der Crash Recovery	697
22.1	Einfuhrung	697
22.2	Organisation eines Data-Managers	699
22.3	Der Recovery-Manager	701
22.4	Ein Undo/Redo-Protokoll	705
22.5	Alternativen zum Undo/Redo-Protokoll	707
22.6	Recovery-Tuning	708
22.7	-Bibliographische Hinweise	709
22.8	ubungen	709

Inhalt	xm
23 Verteilte Transaktionsverarbeitung	711
23.1 Homogene Verteilung	711
23.1.1 Verteilte Transaktionen und Schedules	711
23.1.2 Verteiltes Concurrency Control	715
23.1.3 Verteilte Recovery-Protokolle	720
23.1.4 Transaktionen auf replikativen Datenbanken	728
23.2 Heterogene Verteilung, System-Föderationen	733
23.2.1 Einführung	734
23.2.2 Ein MDBS-Modell	734
23.2.3 Transaktionen und Schedules in einem MDBS	736
23.2.4 Globale Serialisierbarkeit	739
23.2.5 Gewährleistung globaler Serialisierbarkeit	742
23.3 Bibliographische Hinweise	745
23.4 Übungen	746
VI Neuere Entwicklungen	749
24 Ausblicke	751
24.1 Temporale Datenbanken	751
24.2 Probabilistische Datenbanken	755
24.3 Räumliche Daten und GIS	757
24.4 Datenströme (Data Streams)	760
24.5 Data Provenance	762
24.6 Column Stores	764
24.7 Bibliographische Hinweise und Ergänzungen	764
Literaturverzeichnis	767
Index	808