

Inhaltsverzeichnis

Vorwort	17
1 Einleitung und Übersicht	19
1.1 Motivation für den Einsatz eines DBMS	19
1.2 Datenabstraktion	21
1.3 Datenunabhängigkeit	22
1.4 Datenmodelle	23
1.5 Datenbankschema und Ausprägung	24
1.6 Einordnung der Datenmodelle	24
1.6.1 Modelle des konzeptuellen Entwurfs	24
1.6.2 Logische (Implementations-)Datenmodelle	25
1.7 Architekturübersicht eines DBMS	28
1.8 Übungen	30
1.9 Literatur	30
2 Datenbankentwurf	31
2.1 Abstraktionsebenen des Datenbankentwurfs	31
2.2 Allgemeine Entwurfsmethodik	32
2.3 Die Datenbankentwurfsschritte	33
2.4 Die Anforderungsanalyse	33
2.4.1 Informationsstrukturanforderungen	35
2.4.2 Datenverarbeitungsanforderungen	37
2.5 Grundlagen des Entity-Relationship-Modells	37
2.6 Schlüssel	39
2.7 Charakterisierung von Beziehungstypen	39
2.7.1 Funktionalitäten der Beziehungen	39
2.7.2 Funktionalitätsangaben bei n -stelligen Beziehungen	41
2.7.3 Die (min, max) -Notation	44
2.8 Existenzabhängige Entitytypen	48
2.9 Generalisierung	49
2.10 Aggregation	50
2.11 Kombination von Generalisierung und Aggregation	52
2.12 Konsolidierung, Sichtenintegration	53
2.13 Konzeptuelle Modellierung mit UML	59
2.13.1 UML-Klassen	59
2.13.2 Assoziationen zwischen Klassen	60
2.13.3 Aggregation in UML	61
2.13.4 Anwendungsbeispiel: Begrenzungsflächendarstellung von Polyedern in UML	62
2.13.5 Generalisierung in UML-Notation	63
2.13.6 Die Modellierung der Universität in UML	63
2.13.7 Verhaltensmodellierung in UML	64
2.13.8 Anwendungsfall-Modellierung (use cases)	64

2.13.9 Interaktionsdiagramme	66
2.13.10 Interaktionsdiagramm zur Prüfungsdurchführung	66
2.14 Übungen	67
2.15 Literatur	69
3 Das relationale Modell	71
3.1 Definition des relationalen Modells	71
3.1.1 Mathematischer Formalismus	71
3.1.2 Schema-Definition	72
3.2 Umsetzung eines konzeptuellen Schemas in ein relationales Schema	73
3.2.1 Relationale Darstellung von Entitytypen	73
3.2.2 Relationale Darstellung von Beziehungen	73
3.3 Verfeinerung des relationalen Schemas	78
3.3.1 1:N-Beziehungen	78
3.3.2 1:1-Beziehungen	80
3.3.3 Vermeidung von Null-Werten	81
3.3.4 Relationale Modellierung der Generalisierung	82
3.3.5 Beispielausprägung der Universitäts-Datenbank	83
3.3.6 Relationale Modellierung schwacher Entitytypen	85
3.4 Die relationale Algebra	85
3.4.1 Selektion	86
3.4.2 Projektion	87
3.4.3 Vereinigung	87
3.4.4 Mengendifferenz	88
3.4.5 Kartesisches Produkt (Kreuzprodukt)	88
3.4.6 Umbenennung von Relationen und Attributen	89
3.4.7 Definition der relationalen Algebra	90
3.4.8 Der relationale Verbund (Join)	90
3.4.9 Mengendurchschnitt	95
3.4.10 Die relationale Division	96
3.4.11 Operatorbaum-Darstellung	97
3.5 Der Relationenkalkül	97
3.5.1 Beispielanfrage im relationalen Tupelkalkül	98
3.5.2 Quantifizierung von Tupelvariablen	99
3.5.3 Formale Definition des Tupelkalküls	100
3.5.4 Sichere Ausdrücke des Tupelkalküls	101
3.5.5 Der relationale Domänenkalkül	101
3.5.6 Beispielanfragen im Domänenkalkül	102
3.5.7 Sichere Ausdrücke des Domänenkalküls	103
3.6 Ausdruckskraft der Anfragesprachen	104
3.7 Übungen	104
3.8 Literatur	107
4 Relationale Anfragesprachen	109
4.1 Geschichte	109
4.2 Datentypen	110
4.3 Schemadefinition	110
4.4 Schemaveränderung	111
4.5 Elementare Datenmanipulation: Einfügen von Tupeln	112
4.6 Einfache SQL-Anfragen	112

4.7	Anfragen über mehrere Relationen	113
4.8	Aggregatfunktionen und Gruppierung	116
4.9	Geschachtelte Anfragen	117
4.10	Quantifizierte Anfragen in SQL	122
4.11	Nullwerte	124
4.12	Spezielle Sprachkonstrukte	125
4.13	Joins in SQL-92	127
4.14	Rekursion	127
4.15	Veränderungen am Datenbestand	132
4.16	Sichten	134
4.17	Sichten zur Modellierung von Generalisierungen	135
4.18	Charakterisierung update-fähiger Sichten	137
4.19	Einbettung von SQL in Wirtssprachen	138
4.20	Anfragen in Anwendungsprogrammen	139
4.21	JDBC: Java Database Connectivity	142
4.21.1	Verbindungsaufbau zu einer Datenbank	143
4.21.2	Resultset-Programmbeispiel	145
4.21.3	Vorübersetzung von SQL-Ausdrücken	147
4.22	SQLJ: Eine Einbettung von SQL in Java	147
4.23	Query by Example	150
4.24	Übungen	152
4.25	Literatur	156
5	Datenintegrität	157
5.1	Referentielle Integrität	158
5.2	Gewährleistung referentieller Integrität	158
5.3	Referentielle Integrität in SQL	159
5.4	Überprüfung statischer Integritätsbedingungen	160
5.5	Das Universitätsschema mit Integritätsbedingungen	162
5.6	Komplexere Integritätsbedingungen	164
5.7	Trigger	165
5.8	Übungen	167
5.9	Literatur	168
6	Relationale Entwurfstheorie	169
6.1	Funktionale Abhängigkeiten	169
6.1.1	Konventionen zur Notation	170
6.1.2	Einhaltung einer funktionalen Abhängigkeit	170
6.2	Schlüssel	171
6.3	Bestimmung funktionaler Abhängigkeiten	172
6.3.1	Kanonische Überdeckung	175
6.4	„Schlechte“ Relationenschemata	176
6.4.1	Die Updateanomalien	176
6.4.2	Einfügeanomalien	177
6.4.3	Löschanomalien	177
6.5	Zerlegung (Dekomposition) von Relationen	177
6.5.1	Verlustlosigkeit	178
6.5.2	Kriterien für die Verlustlosigkeit einer Zerlegung	180
6.5.3	Abhängigkeitsbewahrung	181
6.6	Erste Normalform	183

6.7	Zweite Normalform	184
6.8	Dritte Normalform	186
6.9	Boyce-Codd Normalform	188
6.10	Mehrwertige Abhängigkeiten	191
6.11	Vierte Normalform	193
6.12	Zusammenfassung	195
6.13	Übungen	196
6.14	Literatur	199
7	Physische Datenorganisation	201
7.1	Speichermedien	201
7.2	Speicherhierarchie	202
7.3	Speicherarrays: RAID	204
7.4	Der Datenbankpuffer	208
7.5	Abbildung von Relationen auf den Sekundärspeicher	209
7.6	Indexstrukturen	211
7.7	ISAM	212
7.8	B-Bäume	214
7.9	B ⁺ -Bäume	218
7.10	Präfix-B ⁺ -Bäume	220
7.11	Hashing	220
7.12	Erweiterbares Hashing	222
7.13	Mehrdimensionale Indexstrukturen	224
7.14	Ballung logisch verwandter Datensätze	229
7.15	Unterstützung eines Anwendungsverhaltens	230
7.16	Physische Datenorganisation in SQL	233
7.17	Übungen	233
7.18	Literatur	235
8	Anfragebearbeitung	237
8.1	Logische Optimierung	238
8.1.1	Äquivalenzen in der relationalen Algebra	240
8.1.2	Anwendung der Transformationsregeln	242
8.2	Physische Optimierung	246
8.2.1	Implementierung der Selektion	248
8.2.2	Implementierung von binären Zuordnungsoperatoren	248
8.2.3	Gruppierung und Duplikateliminierung	255
8.2.4	Projektion und Vereinigung	256
8.2.5	Zwischenspeicherung	256
8.2.6	Übersetzung der logischen Algebra	258
8.3	Kostenmodelle	261
8.3.1	Selektivitäten	262
8.3.2	Kostenabschätzung für die Selektion	264
8.3.3	Kostenabschätzung für den Join	265
8.3.4	Kostenabschätzung für die Sortierung	265
8.4	„Tuning“ von Datenbankankfragen	266
8.5	Kostenbasierte Optimierer	267
8.5.1	Suchraum für die Join-Optimierung	268
8.5.2	Dynamische Programmierung	270
8.6	Übungen	274

8.7	Literatur	276
9	Transaktionsverwaltung	279
9.1	Begriffsbildung	279
9.2	Anforderungen an die Transaktionsverwaltung	280
9.3	Operationen auf Transaktions-Ebene	280
9.4	Abschluss einer Transaktion	281
9.5	Eigenschaften von Transaktionen	283
9.6	Transaktionsverwaltung in SQL	284
9.7	Zustandsübergänge einer Transaktion	285
9.8	Literatur	286
10	Fehlerbehandlung	287
10.1	Fehlerklassifikation	287
10.1.1	Lokaler Fehler einer Transaktion	287
10.1.2	Fehler mit Hauptspeicherverlust	288
10.1.3	Fehler mit Hintergrundspeicherverlust	289
10.2	Die Speicherhierarchie	289
10.2.1	Ersetzung von Puffer-Seiten	289
10.2.2	Einbringen von Änderungen einer Transaktion	290
10.2.3	Einbringstrategie	291
10.2.4	Hier zugrunde gelegte Systemkonfiguration	292
10.3	Protokollierung von Änderungsoperationen	292
10.3.1	Struktur der Log-Einträge	293
10.3.2	Beispiel einer Log-Datei	293
10.3.3	Logische oder physische Protokollierung	293
10.3.4	Schreiben der Log-Information	294
10.3.5	Das WAL-Prinzip	296
10.4	Wiederanlauf nach einem Fehler	296
10.4.1	Analyse des Logs	297
10.4.2	Redo-Phase	298
10.4.3	Undo-Phase	298
10.5	Fehlertoleranz des Wiederanlaufs	298
10.6	Lokales Zurücksetzen einer Transaktion	300
10.7	Partielles Zurücksetzen einer Transaktion	301
10.8	Sicherungspunkte	302
10.8.1	Transaktionskonsistente Sicherungspunkte	302
10.8.2	Aktionskonsistente Sicherungspunkte	303
10.8.3	Unschärfe (fuzzy) Sicherungspunkte	305
10.9	Recovery nach einem Verlust der materialisierten Datenbasis	306
10.10	Übungen	307
10.11	Literatur	308
11	Mehrbenutzersynchronisation	309
11.1	Fehler bei unkontrolliertem Mehrbenutzerbetrieb	310
11.1.1	Verlorengegangene Änderungen (<i>lost update</i>)	310
11.1.2	Abhängigkeit von nicht freigegebenen Änderungen	310
11.1.3	Phantomproblem	311
11.2	Serialisierbarkeit	311
11.2.1	Beispiele serialisierbarer Ausführungen (Historien)	312

11.2.2	Nicht serialisierbare Historie	312
11.3	Theorie der Serialisierbarkeit	315
11.3.1	Definition einer Transaktion	315
11.3.2	Historie (Schedule)	316
11.3.3	Äquivalenz zweier Historien	317
11.3.4	Serialisierbare Historien	318
11.3.5	Kriterien für Serialisierbarkeit	318
11.4	Eigenschaften von Historien bezüglich der Recovery	320
11.4.1	Rücksetzbare Historien	320
11.4.2	Historien ohne kaskadierendes Rücksetzen	320
11.4.3	Strikte Historien	321
11.4.4	Beziehungen zwischen den Klassen von Historien	321
11.5	Der Datenbank-Scheduler	322
11.6	Sperrbasierte Synchronisation	323
11.6.1	Zwei Sperrmodi	323
11.6.2	Zwei-Phasen-Sperrprotokoll	324
11.6.3	Kaskadierendes Rücksetzen (Schneeballeffekt)	326
11.7	Verklemmungen (Deadlocks)	326
11.7.1	Erkennung von Verklemmungen	327
11.7.2	Preclaiming zur Vermeidung von Verklemmungen	328
11.7.3	Verklemmungsvermeidung durch Zeitstempel	329
11.8	Hierarchische Sperrgranulate	330
11.9	Einfüge- und Löschoptionen, Phantome	334
11.10	Zeitstempel-basierende Synchronisation	335
11.11	Optimistische Synchronisation	337
11.12	Klassifizierung der Verfahren	338
11.13	Synchronisation von Indexstrukturen	338
11.14	Mehrbenutzersynchronisation in SQL-92	342
11.15	Übungen	344
11.16	Literatur	347
12	Sicherheitsaspekte	349
12.1	Discretionary Access Control	351
12.2	Zugriffskontrolle in SQL	351
12.2.1	Identifikation und Authentisierung	352
12.2.2	Autorisierung und Zugriffskontrolle	352
12.2.3	Sichten	353
12.2.4	Individuelle Sicht für eine Benutzergruppe	354
12.2.5	Auditing	355
12.3	Verfeinerung des Autorisierungsmodells	355
12.3.1	Rollenbasierte Autorisierung: Implizite Autorisierung von Subjekten	356
12.3.2	Implizite Autorisierung von Operationen	358
12.3.3	Implizite Autorisierung von Objekten	358
12.3.4	Implizite Autorisierung entlang einer Typhierarchie	359
12.4	Mandatory Access Control	360
12.5	Multilevel-Datenbanken	361
12.6	Kryptographie	364
12.6.1	Der Data Encryption Standard	365

12.6.2	Public-Key Kryptographie	366
12.6.3	Public Key Infrastruktur (PKI)	368
12.7	Zusammenfassung	368
12.8	Übungen	369
12.9	Literatur	370
13	Objektorientierte Datenbanken	371
13.1	Bestandsaufnahme relationaler Datenbanksysteme	371
13.2	Vorteile der objektorientierten Datenmodellierung	375
13.3	Der ODMG-Standard	376
13.4	Eigenschaften von Objekten	377
13.4.1	Objektidentität	378
13.4.2	Typ eines Objekts	379
13.4.3	Wert eines Objekts	379
13.5	Definition von Objekttypen	380
13.5.1	Attribute	380
13.5.2	Beziehungen	380
13.5.3	Typeigenschaften: Extensionen und Schlüssel	387
13.6	Modellierung des Verhaltens: Operationen	387
13.7	Vererbung und Subtypisierung	390
13.7.1	Terminologie	390
13.7.2	Einfache und Mehrfachvererbung	391
13.8	Beispiel einer Typhierarchie	392
13.9	Verfeinerung (Spezialisierung) und spätes Binden von Operationen	395
13.10	Mehrfachvererbung	398
13.11	Die Anfragesprache OQL	399
13.11.1	Einfache Anfragen	399
13.11.2	Geschachtelte Anfragen und Partitionierung	400
13.11.3	Pfadausdrücke	401
13.11.4	Erzeugung von Objekten	402
13.11.5	Operationsaufruf	402
13.12	C++-Einbettung	402
13.12.1	Objektidentität	404
13.12.2	Objekterzeugung und Ballung	405
13.12.3	Einbettung von Anfragen	405
13.13	Übungen	406
13.14	Literatur	407
14	Erweiterbare und objekt-relationale Datenbanken	409
14.1	Übersicht über die objekt-relationalen Konzepte	409
14.2	Large Objects (LOBs)	410
14.3	Distinct Types: Einfache benutzerdefinierte Datentypen	412
14.4	Table Functions	416
14.4.1	Nutzung einer <i>Table Function</i> in Anfragen	417
14.4.2	Implementierung einer <i>Table Function</i>	417
14.5	Benutzerdefinierte strukturierte Objekttypen	419
14.6	Geschachtelte Objekt-Relationen	423
14.7	Vererbung von SQL-Objekttypen	427
14.8	Komplexe Attribut-Typen	430
14.9	Übungen	431

14.10 Literatur	432
15 Deduktive Datenbanken	433
15.1 Terminologie	433
15.2 Datalog	433
15.3 Eigenschaften von Datalog-Programmen	437
15.3.1 Rekursivität	437
15.3.2 Sicherheit von Datalog-Regeln	437
15.4 Auswertung von nicht-rekursiven Datalog-Programmen	438
15.4.1 Auswertung eines Beispielprogramms	438
15.4.2 Auswertungs-Algorithmus	441
15.5 Auswertung rekursiver Regeln	443
15.6 Inkrementelle (semi-naive) Auswertung rekursiver Regeln	445
15.7 Bottom-Up oder Top-Down Auswertung	449
15.8 Negation im Regelrumpf	451
15.8.1 Stratifizierte Datalog-Programme	451
15.8.2 Auswertung von Regeln mit Negation	452
15.8.3 Ein etwas komplexeres Beispiel	453
15.9 Ausdruckskraft von Datalog	453
15.10 Übungen	455
15.11 Literatur	459
16 Verteilte Datenbanken	461
16.1 Terminologie und Abgrenzung	461
16.2 Entwurf verteilter Datenbanken	463
16.3 Horizontale und vertikale Fragmentierung	465
16.3.1 Horizontale Fragmentierung	466
16.3.2 Abgeleitete horizontale Fragmentierung	468
16.3.3 Vertikale Fragmentierung	469
16.3.4 Kombinierte Fragmentierung	471
16.3.5 Allokation für unser Beispiel	472
16.4 Transparenz in verteilten Datenbanken	473
16.4.1 Fragmentierungstransparenz	473
16.4.2 Allokationstransparenz	474
16.4.3 Lokale Schema-Transparenz	474
16.5 Anfrageübersetzung und -optimierung in VDBMS	475
16.5.1 Anfragebearbeitung bei horizontaler Fragmentierung	475
16.5.2 Anfragebearbeitung bei vertikaler Fragmentierung	477
16.6 Join-Auswertung in VDBMS	479
16.6.1 Join-Auswertung ohne Filterung	479
16.6.2 Join-Auswertung mit Semijoin-Filterung	480
16.6.3 Join-Auswertung mit Bitmap-Filterung	482
16.7 Transaktionskontrolle in VDBMS	484
16.8 Mehrbenutzersynchronisation in VDBMS	489
16.8.1 Serialisierbarkeit	489
16.8.2 Das Zwei-Phasen-Sperrprotokoll in VDBMS	489
16.9 Deadlocks in VDBMS	490
16.9.1 Erkennung von Deadlocks	490
16.9.2 Deadlock-Vermeidung	493
16.10 Synchronisation bei replizierten Daten	494

16.11	Peer-to-Peer-Informationssysteme	497
16.11.1	P2P-Systeme für den Datenaustausch (File-Sharing)	497
16.11.2	Verteilte Hashtabellen (Distributed Hash Tables DHTs)	499
16.11.3	Mehrdimensionaler P2P-Datenraum	503
16.12	Übungen	505
16.13	Literatur	509
17	Betriebliche Anwendungen: OLTP, Data Warehouse, Data Mining	511
17.1	SAP R/3: Ein betriebswirtschaftliches Datenbankanwendungssystem	511
17.1.1	Architektur von SAP R/3	511
17.1.2	Datenmodell und Schema von SAP R/3	512
17.1.3	ABAP/4	513
17.1.4	Transaktionen in SAP R/3	516
17.2	Data Warehouse, Decision-Support, OLAP	517
17.2.1	Datenbankentwurf für das Data Warehouse	518
17.2.2	Anfragen im Sternschema: Star Join	521
17.2.3	Roll-Up/Drill-Down-Anfragen	522
17.2.4	Flexible Auswertungsmethoden	524
17.2.5	Materialisierung von Aggregaten	524
17.2.6	Der cube -Operator	526
17.2.7	Wiederverwendung materialisierter Aggregate	526
17.2.8	Bitmap-Indices für OLAP-Anfragen	529
17.2.9	Auswertungsalgorithmen für komplexe OLAP-Anfragen	530
17.3	Bewertung (Ranking) von Objekten	532
17.3.1	Top-k-Anfragen	533
17.3.2	Skyline-Anfragen	537
17.3.3	Data Warehouse-Architekturen	540
17.3.4	Attribut-basierte Speicherung (Column-Stores)	541
17.4	Data Mining	543
17.4.1	Klassifikation von Objekten	543
17.4.2	Assoziationsregeln	544
17.4.3	Der Ä Priori-Algorithmus	545
17.4.4	Bestimmung der Assoziationsregeln	547
17.4.5	Cluster-Bestimmung	548
17.5	Übungen	549
17.6	Literatur	550
18	Internet-Datenbankanbindungen	553
18.1	HTML- und HTTP-Grundlagen	553
18.1.1	HTML: Die Hypertext-Sprache des World Wide Web	553
18.1.2	Adressierung von Web-Dokumenten	554
18.1.3	Client/Server-Architektur des World Wide Web	556
18.1.4	HTTP: Das HyperText Transfer Protokoll	556
18.1.5	HTTPS	557
18.2	Web-Datenbank-Anbindung via Servlets	558
18.2.1	Beispiel-Servlet	558
18.3	Java Server Pages / Active Server Pages	564
18.3.1	JSP/HTML-Seite mit Java-Code	565
18.3.2	HTML-Seite mit Java-Bean-Aufruf	567
18.3.3	Die Java-Bean Komponente <i>VorlesungenBean</i>	568

18.3.4	Sokrates' Homepage	570
18.4	Datenbankanbindung via Java-Applets	570
18.5	Übungen	571
18.6	Literatur	572
19	XML-Datenmodellierung und Web-Services	573
19.1	XML-Datenmodellierung	573
19.1.1	Schema oder kein Schema	574
19.1.2	Rekursive Schemata	576
19.1.3	Universitätsinformation in XML-Format	576
19.1.4	XML-Namensräume	578
19.1.5	XML Schema: Eine Schemadefinitionssprache	580
19.1.6	Verweise (Referenzen) in XML-Daten	582
19.2	XQuery: Eine XML-Anfragesprache	583
19.2.1	Pfadausdrücke	583
19.2.2	Verkürzte XPath-Syntax	588
19.2.3	Beispiel-Pfadausdrücke in verkürzter Syntax	589
19.2.4	Anfragesyntax von XQuery	590
19.2.5	Geschachtelte Anfragen	592
19.2.6	Joins in XQuery	592
19.2.7	Join-Prädikat im Pfadausdruck	593
19.2.8	Das let-Konstrukt	594
19.2.9	Dereferenzierung in FLWOR-Ausdrücken	595
19.2.10	Das if-then-else-Konstrukt	597
19.2.11	Rekursive Anfragen	598
19.3	Zusammenspiel von relationalen Datenbanken und XML	600
19.3.1	XML-Repräsentation gemäß Pre- und Postorder-Rängen	606
19.3.2	Der neue Datentyp xml	610
19.3.3	Änderungen der XML-Dokumente	614
19.3.4	Publikation relationaler Daten als XML-Dokumente	615
19.3.5	Fallstudie: XML-Unterstützung in IBM DB2 V9	619
19.4	Web-Services	624
19.4.1	Erstellen und Nutzen eines Web-Services im Überblick	626
19.4.2	Das Auffinden von Diensten	628
19.4.3	Ein Beispiel-Web-Service	630
19.4.4	Definition der Web-Service-Schnittstellen	630
19.4.5	Nachrichtenformat für die Interaktion mit Web-Services	633
19.4.6	Implementierung des Web-Services	635
19.4.7	Aufruf des Web-Services	636
19.5	Übungen	638
19.6	Literatur	641
20	Leistungsbewertung	645
20.1	Überblick über Datenbanksystem-Benchmarks	645
20.2	Der TPC-C Benchmark	645
20.3	Die TPC-H und TPC-R (früher TPC-D) Benchmarks	648
20.4	Der OO7 Benchmark für oo-Datenbanken	654
20.5	Der TPC-W Benchmark	655
20.6	Neue TPC-Benchmarks	658
20.6.1	TPC-E: Der neue OLTP-Benchmark	658

20.6.2 TPC-App: der neue Webservice-Benchmark	659
20.6.3 TPC-DS: der neue Decision Support Benchmark	660
20.7 Übungen	661
20.8 Literatur	662
Literaturverzeichnis	663
Index	702