

Thomas Kudraß

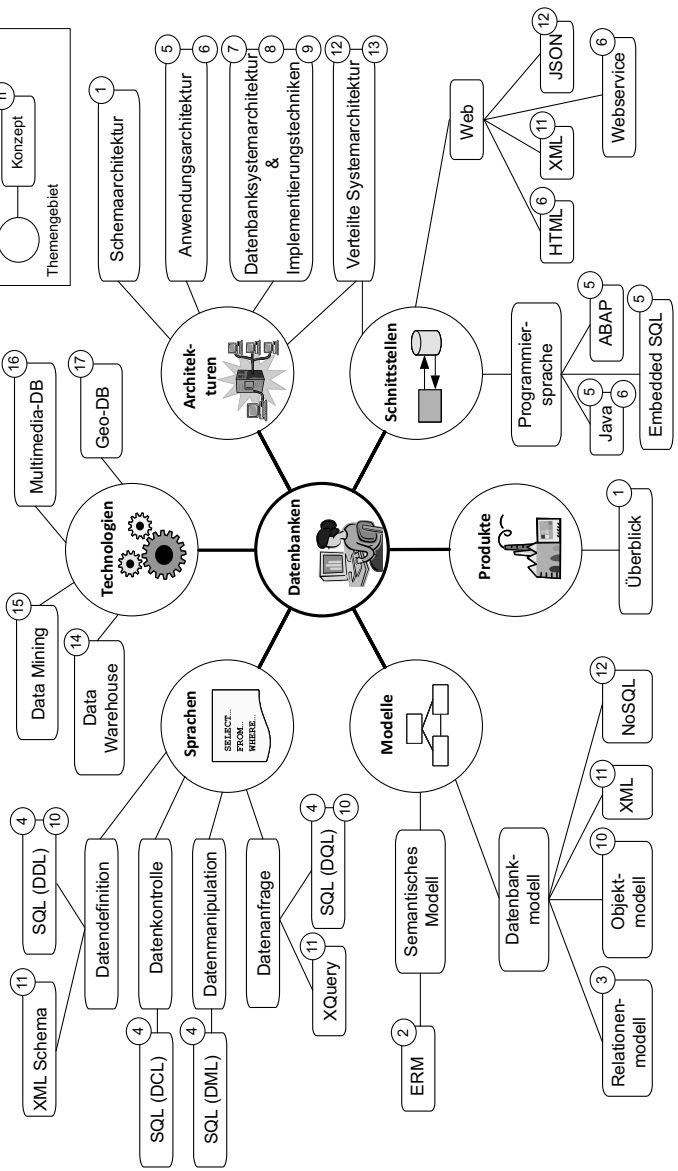
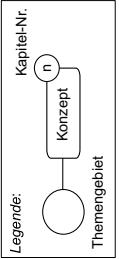
Taschenbuch Datenbanken



2. Auflage



HANSER



1	Datenbanken: Grundlagen und Überblick	19
2	Informationsmodellierung	44
3	Relationales Datenmodell	67
4	Die Datenbanksprache SQL	99
5	Datenbank-Anwendungsprogrammierung	144
6	Datenbanken im Web	189
7	Komponenten eines Datenbankmanagementsystems	227
8	Dateiorganisation und Indexe	269
9	Optimierung von Datenbanken und Leistungsbewertung	287
10	Objektrelationale Datenbanken	322
11	XML und Datenbanken	345
12	NoSQL-Datenbanksysteme	368
13	Verteilte und föderierte Datenbanksysteme	394
14	Data Warehouse	430
15	Data Mining	461
16	Multimedia-Datenbanken	478
17	Geodatenbanken	499

Kudraß
Taschenbuch Datenbanken

Bleiben Sie auf dem Laufenden!



Unser **Computerbuch-Newsletter** informiert Sie monatlich über neue Bücher und Termine. Profitieren Sie auch von Gewinnspielen und exklusiven Leseproben. Gleich anmelden unter



www.hanser-fachbuch.de/newsletter

Herausgeber

Prof. Dr. Thomas Kudraß, Hochschule für Technik, Wirtschaft und Kultur Leipzig

Autoren

Prof. Dr. Thomas Brinkhoff

Jade Hochschule Wilhelmshaven/Oldenburg/Elsfleth

Prof. Dr. Olaf Herden

Duale Hochschule Baden-Württemberg Stuttgart, Campus Horb

Prof. Dr. Harm Knolle

Hochschule Bonn-Rhein-Sieg

Prof. Dr. Thomas Kudraß

Hochschule für Technik, Wirtschaft und Kultur Leipzig

Prof. Dr. Klaus Meyer-Wegener

Friedrich-Alexander-Universität Erlangen-Nürnberg

Prof. Dr. Thomas Rakow

Fachhochschule Düsseldorf

Prof. Dr. Norbert Ritter

Universität Hamburg

Prof. Dr. Kai-Uwe Sattler

Technische Universität Ilmenau

Prof. Dr. Petra Sauer

Beuth Hochschule für Technik Berlin

Prof. Dr. Bernhard Schiefer

Hochschule Kaiserslautern, Campus Zweibrücken

Dr. Harald Schöning

Software AG Darmstadt

Prof. Dr. Uta Störl

Hochschule Darmstadt

Dr. Can Türker

Functional Genomics Center Zürich

Taschenbuch Datenbanken

2., neu bearbeitete Auflage

herausgegeben von

Thomas Kudraß

Mit 163 Bildern und 30 Tabellen



Fachbuchverlag Leipzig
im Carl Hanser Verlag



Alle in diesem Buch enthaltenen Programme, Verfahren und elektronischen Schaltungen wurden nach bestem Wissen erstellt und mit Sorgfalt getestet. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund ist das im vorliegenden Buch enthaltene Programm-Material mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor und Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Art aus der Benutzung dieses Programm-Materials oder Teilen davon entsteht.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

ISBN 978-3-446-43508-7

E-Book-ISBN 978-3-446-44026-5

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdruckes und der Vervielfältigung des Buches, oder Teilen daraus, vorbehalten. Kein Teil des Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder ein anderes Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung – mit Ausnahme der in den §§ 53, 54 URG genannten Sonderfälle –, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

© 2015 Carl Hanser Verlag München
Internet: <http://www.hanser-fachbuch.de>

Lektorat: Franziska Jacob, M. A.

Herstellung: Der Buchmacher, Arthur Lenner, München

Satz: Satzherstellung Dr. Naake, Brand-Erbisdorf

Coverrealisierung: Stephan Rönigk

Druck und Bindung: Kösel, Krugzell

Printed in Germany

Vorwort

Informationen sind der Rohstoff des Informationszeitalters, in dem wir heute leben. Der größte Teil davon wird auf digitalen Medien gespeichert, wobei Datenbanksysteme zur Verwaltung der Inhalte eine zentrale Rolle spielen. Datenbanken bilden quasi das Rückgrat von Informationssystemen in allen Bereichen – oft gar nicht sichtbar für den Endanwender. So ist es nicht verwunderlich, dass sich das Fach Datenbanken in den letzten Jahren als Kerndisziplin der Informatik etabliert hat, zugleich aber auch immer stärker in die Ausbildung anderer Berufe eingeflossen ist.

Das Taschenbuch Datenbanken erscheint in der bekannten Taschenbuchreihe des Fachbuchverlags Leipzig, von dem auch die Anregung für ein solches Buch kam. Das gesamte Buch ist als Nachschlagewerk für das heutzutage etablierte Wissen im Fach Datenbanken konzipiert. Grundlegende Fakten und Zusammenhänge werden in kompakter und übersichtlicher Form dargestellt. Alle Begriffe sind darüber hinaus in einem umfangreichen Sachwortverzeichnis zu finden. An vielen Stellen gibt es Querverweise und Empfehlungen zu weiterführender Literatur.

Im Taschenbuch werden die wichtigsten theoretischen Grundlagen von Datenbanken behandelt, zugleich aber auch eine Vielzahl von Technologien, die Bestandteil moderner Datenbanksysteme sind. Angesichts des begrenzten Umfangs konnten dabei nicht alle Themen berücksichtigt werden, insbesondere wenn sie noch Gegenstand der Forschung sind.

Das Buch ist in 17 Kapitel gegliedert. Die Kapitel 1 bis 6 dienen vor allem dem Einstieg in das Gebiet und beinhalten Grundlagenwissen über den Datenbankentwurf und die Entwicklung von Datenbankanwendungen. Dabei wird der Datenbanksprache SQL breiter Raum eingeräumt, ebenso der Verbindung von Web und Datenbanken. Die Kapitel 7 bis 9 behandeln die interne Arbeitsweise von Datenbanksystemen, deren Verständnis für die Datenbankoptimierung und -administration notwendig ist. Die Kapitel 10 bis 13 beinhalten Basistechnologien, die für den fortgeschrittenen Datenbanknutzer interessant sind, z. B. objektrelationale und XML-Datenbanken. Zunehmende Bedeutung in Unternehmen erlangt auch die Verteilung und Integration von Datenbanken, so dass diesem Thema ein gesondertes Kapitel gewidmet ist. Die Kapitel 14 bis 17 behandeln Datenbanktechnologien, wie sie in speziellen Informationssystemen genutzt werden. So werden Data Warehouse und Data Mining als Basis entscheidungsunterstützender Systeme dargestellt. Multimediale Daten und Geodaten unterscheiden sich

gegenüber „klassischen“ Daten und werden deshalb in eigenen Kapiteln beschrieben.

Die Idee des Buches besteht darin, vor allem allgemeingültige Konzepte und Technologien darzustellen, nicht aber Produkte einzelner Anbieter. Aufgrund der weitgehenden Standardisierung von Sprachen und Schnittstellen bildet das im Buch präsentierte Wissen eine solide Grundlage zum Verständnis herstellerepezifischer Systeme. Somit eignet es sich für jeden an Datenbanken interessierten Leser: ob Student, Praktiker oder „Quereinsteiger“. Für weiterführende Informationen und Aktualisierungen sei noch auf die zugehörige Webseite verwiesen.

Als Herausgeber möchte ich mich ganz herzlich bei allen Autoren bedanken, die zu den einzelnen Kapiteln ihr spezielles Fachwissen beigetragen haben. Mein besonderer Dank geht an Erika Hotho und Franziska Kaufmann vom Fachbuchverlag Leipzig für die hervorragende Betreuung und die stets angenehme Zusammenarbeit, ebenso an PTP-Berlin für das gelungene Layout. Danken möchte ich auch Hans-Peter Leidhold und Dörte König für Korrekturhinweise zum Manuskript. Gleichfalls ein Dank an Pieter Hauffe für seine Unterstützung bei der technischen Vorbereitung des Buchprojekts.

Ich hoffe, dass das Buch für den Leser zu einer wertvollen Hilfe im Begriffsdschungel der Datenbankwelt wird. Anregungen, Verbesserungsvorschläge oder Kritik werden jederzeit gern entgegengenommen.

Leipzig, im August 2007

Thomas Kudraß

Vorwort zur 2. Auflage

In dieser 2. Auflage wurden die Ausführungen überarbeitet und aktuelle Entwicklungen aus dem Bereich „Big Data“ aufgegriffen. NoSQL-Datenbanken, Column Stores in Data-Warehouse-Systemen und Cloud-Datenbanken sind dabei einige der neuen Themen, die ins Buch aufgenommen wurden. Ein Dank gilt den Rezensenten der Erstauflage, von denen zahlreiche Hinweise berücksichtigt wurden. Ebenso möchte ich mich herzlich für die angenehme Zusammenarbeit bedanken: bei Franziska Jacob vom Fachbuchverlag Leipzig, bei Arthur Lenner und Dr. Steffen Naake, die für Lektorat, Herstellung und Satzlegung verantwortlich waren. Kommentare zu dieser 2. Auflage werden wiederum gern entgegengenommen.

Leipzig, im Januar 2015

Thomas Kudraß

Inhaltsverzeichnis

1	Datenbanken: Grundlagen und Überblick	19
1.1	<i>Dateien und Dateisysteme</i>	19
1.2	<i>Terminologie</i>	20
1.3	<i>Merkmale eines DBMS</i>	23
1.3.1	Aufgaben eines DBMS	23
1.3.2	Vorteile des Datenbankeinsatzes	24
1.3.3	Nachteile von Datenbanksystemen	25
1.3.4	Produkte	25
1.4	<i>Architektur eines Datenbanksystems</i>	26
1.4.1	Architekturen	26
1.4.2	Schemaarchitektur	27
1.4.2.1	Datenbankschema	27
1.4.2.2	Drei-Ebenen-Architektur	28
1.4.3	Datenunabhängigkeit	29
1.5	<i>Benutzerrollen bei Entwurf und Betrieb von Datenbanken</i>	30
1.6	<i>Datenbanken und Informationssysteme</i>	32
1.7	<i>Fachgebiet Datenbanken im Überblick</i>	33
1.7.1	Themenbereiche und Zusammenhänge	33
1.7.2	Einordnung des Fachs innerhalb der Informatik	35
1.8	<i>Historische Entwicklung</i>	37
1.8.1	Frühzeit	38
1.8.2	Prärelationale DBMS	38
1.8.3	Durchbruch der relationalen Datenbanken	39
1.8.4	Neue Anwendungsfelder für Datenbanken	40
1.8.5	Neuzeit	40
1.9	<i>Datenbanken in der Forschung</i>	41
1.9.1	Fachverbände	41
1.9.2	Aktuelle Forschungstrends	42
2	Informationsmodellierung	44
2.1	<i>Datenbankentwurf</i>	44
2.1.1	Phasenmodell	44
2.1.2	Anforderungsanalyse	45
2.1.3	Konzeptioneller Entwurf	46
2.1.4	Logischer Entwurf	47
2.1.5	Datendefinition	48
2.1.6	Physischer Entwurf	49
2.1.7	Implementierung und Wartung	49
2.2	<i>Grundlagen des Entity-Relationship-Modells (ERM)</i>	49
2.2.1	Semantische Datenmodelle	50
2.2.2	Grundkonzepte des klassischen ERM	50
2.2.2.1	Konzepte auf der Instanzebene	51

2.2.2.2	Konzepte auf der Typebene	52
2.2.2.3	Rollenmodellierung	53
2.2.3	Kardinalitäten von Beziehungstypen	54
2.2.3.1	Kardinalitäten im klassischen ERM	54
2.2.3.2	Intervalle und Komplexitätsgrade	54
2.2.4	Existenzabhängigkeit vs. Optionalität	55
2.2.5	Rekursive und n -äre Beziehungstypen	56
2.2.5.1	Rekursive Beziehungstypen	56
2.2.5.2	N -äre Beziehungstypen	57
2.2.6	Attribute	58
2.2.7	Modellierungsbeispiel	59
2.3	Erweiterungen des ERM	60
2.3.1	Erweiterungen bei Attributen	60
2.3.2	Generalisierung und Spezialisierung	61
2.3.3	Aggregation	64
2.3.4	Modellierung zeitlicher Aspekte	65
3	Relationales Datenmodell	67
3.1	Konzepte und Grundbegriffe des relationalen Datenmodells	67
3.1.1	Relationen, Tupel, Attribute und Wertebereiche	67
3.1.2	Eigenschaften von Relationen	69
3.2	Integritätsbedingungen	70
3.2.1	Grundbegriffe	70
3.2.2	Entitätsintegrität	71
3.2.3	Referenzielle Integrität	71
3.2.3.1	Begriff	71
3.2.3.2	Regeln für Fremdschlüssel	72
3.2.3.3	Gewährleistung der referenziellen Integrität bei kritischen DML-Operationen	73
3.3	Abbildung des EERM auf das relationale Datenmodell	75
3.3.1	Problemstellung	75
3.3.2	Abbildungsregeln für Attribute und Entitytypen	75
3.3.3	Abbildungsregeln für Beziehungstypen	76
3.3.4	Abbildungsregeln für die Generalisierung	79
3.4	Optimierung von Relationen	80
3.4.1	Problemstellung	80
3.4.2	Anomalien bei DML-Operationen auf Relationen	81
3.4.3	Abhängigkeiten	82
3.4.3.1	Funktionale Abhängigkeiten	82
3.4.3.2	Mehrwertige Abhängigkeiten	83
3.4.4	Verbundtreue und Abhängigkeitstreue	84
3.4.5	Normalformenlehre	85
3.4.5.1	Erste Normalform (1NF)	85
3.4.5.2	Zweite Normalform (2NF)	86
3.4.5.3	Dritte Normalform (3NF)	87
3.4.5.4	Boyce-Codd-Normalform (BCNF)	88
3.4.5.5	Vierte Normalform (4NF)	89

3.4.5.6	Fünfte Normalform (5NF)	90
3.4.5.7	Denormalisierung	92
3.5	<i>Operationen der Relationenalgebra</i>	92
3.5.1	Einführung	92
3.5.2	Relationenorientierte Operationen	93
3.5.3	Mengenoperationen	95
3.5.4	Relationenalgebra und relationale Sprachen	96
3.5.5	Relationenkalkül	97
4	Die Datenbanksprache SQL	99
4.1	<i>Grundkonzepte</i>	99
4.2	<i>Historie</i>	101
4.3	<i>Spezifikationsdokumente</i>	102
4.4	<i>Beispieldatenbank</i>	103
4.5	<i>Datenbankanfragen</i>	103
4.5.1	Einführende Beispielanfragen	104
4.5.2	Grundgerüst von Anfragen	105
4.5.3	Anfragen mit Aggregatfunktionen	107
4.5.4	Anfragen mit Tabellenfunktionen	108
4.5.5	Anfragen mit Mengenoperationen	108
4.5.6	Anfragen mit Verbundoperationen	110
4.5.7	Verschachtelte Anfragen	112
4.5.8	Rekursive Anfragen	114
4.5.9	Anfragen mit sortierter Ausgabe	115
4.6	<i>Datenmanipulation</i>	116
4.6.1	Einfügen von Tabellenzeilen	116
4.6.2	Ändern von Tabellenzeilen	117
4.6.3	Löschen von Tabellenzeilen	118
4.6.4	Zusammenführen von Tabellenzeilen	118
4.7	<i>Datendefinition</i>	119
4.7.1	SQL-Datentypen	119
4.7.2	Erzeugen und Löschen von Schemata	124
4.7.3	Erzeugen, Ändern und Löschen von Tabellen	124
4.7.4	Erzeugen und Löschen von Domänen	127
4.7.5	Erzeugen und Löschen von Integritätsbedingungen	128
4.7.6	Erzeugen und Löschen von Sichten	130
4.7.7	Erzeugen und Löschen von Routinen	131
4.7.7.1	Prozedurale SQL-Konstrukte	131
4.7.7.2	Erzeugen von SQL-Routinen	135
4.7.7.3	Erzeugen von externen Routinen	136
4.7.7.4	Löschen von Routinen	136
4.7.8	Erzeugen und Löschen von Triggern	137
4.7.9	Erzeugen und Löschen von Sequenzgeneratoren	138
4.8	<i>Transaktionssteuerung</i>	138
4.8.1	Transaktionsanweisungen	139
4.8.2	Isolationsebenen	140
4.8.3	Überprüfung von Integritätsbedingungen	141

4.9	Zugriffskontrolle	141
4.9.1	Vergabe von Zugriffsrechten	141
4.9.2	Zurücknahme von Zugriffsrechten	143
4.9.3	Erzeugen und Löschen von Rollen	143
5	Datenbank-Anwendungsprogrammierung	144
5.1	Grundlagen der Datenbank-Anwendungsprogrammierung	144
5.1.1	Impedance Mismatch	144
5.1.2	Einbettungstechniken	145
5.1.3	Einbettungsarten	145
5.1.3.1	Statische Programmierspracheneinbettung	145
5.1.3.2	Dynamische Programmierspracheneinbettung	146
5.1.4	Architekturansätze	147
5.1.4.1	Fat-Client-Architektur	148
5.1.4.2	Thin-Client-Architektur	148
5.1.4.3	Thin-Client-Architektur mit Applikationsserver	150
5.2	Embedded SQL	151
5.2.1	Grundidee und Architektur	151
5.2.2	Syntax	152
5.2.3	Host-Variablen	152
5.2.4	Das Cursor-Konzept	152
5.2.5	Statische Einbettung	154
5.2.6	Dynamische Einbettung	155
5.2.7	Indikatorvariablen	156
5.2.8	SQLJ	156
5.3	Aufrufschnittstellen	158
5.3.1	Überblick	158
5.3.2	SQL/CLI	158
5.3.3	JDBC	160
5.3.3.1	Treibertypen	161
5.3.3.2	Verbindung aufbauen	161
5.3.3.3	Verbindung schließen	163
5.3.3.4	Transaktionen	163
5.3.3.5	Leseanweisungen – das ResultSet	164
5.3.3.6	Änderungen	165
5.3.3.7	Zugriff auf Metadaten	167
5.4	Relationale Datenbankprogrammiersprachen	168
5.4.1	Datenbankinterne Ansätze	168
5.4.1.1	SQL/PSM	168
5.4.1.2	Benutzerdefinierte Routinen	169
5.4.1.3	Trigger	170
5.4.2	4GL-Programmiersprachen	174
5.4.2.1	Überblick	174
5.4.2.2	Beispiel: ABAP	174
5.5	Objektrelationale Abbildung	177
5.5.1	Java Persistence API (JPA)	179
5.5.1.1	Schemaabbildung	180

5.5.1.2	Datenzugriff	181
5.5.1.3	Formulieren von Anfragen	183
5.5.2	JDO – Java Data Objects	184
5.5.3	Entity Beans	188
6	Datenbanken im Web	189
6.1	<i>Grundlagen des Web</i>	189
6.2	<i>Eigenschaften von Webanwendungen</i>	191
6.2.1	Anforderungen	192
6.2.2	Webseiten	192
6.2.3	Dynamische Webseiten	195
6.2.4	Adressierung	197
6.2.5	Kommunikation	198
6.3	<i>Datenbankanbindung im Web</i>	200
6.3.1	Architekturen	201
6.3.2	Programmierung	202
6.3.3	Überblick	204
6.4	<i>Datenbankanbindung über Datenexport</i>	206
6.5	<i>Clientseitige Datenbankanbindung</i>	206
6.5.1	Skripteinbettung	207
6.5.2	Programmeinbettung	209
6.5.3	Weitere Techniken	212
6.6	<i>Serverseitige Datenbankanbindung</i>	212
6.6.1	Webseitengeneratoren	212
6.6.2	Skripteinbettung	215
6.6.3	Programmeinbettung	219
6.6.4	Applikationsserver	221
6.6.5	Webservices	223
7	Komponenten eines Datenbankmanagementsystems	227
7.1	<i>Architektur von DBMS</i>	227
7.1.1	Schichtenmodell	227
7.1.2	Prozessarchitektur	229
7.2	<i>Pufferverwaltung</i>	230
7.2.1	Notwendigkeit und Aufgabe	231
7.2.2	Speicherzuteilung	232
7.2.3	Seitenersetzung	233
7.3	<i>Speicher- und Zugriffssystem</i>	234
7.3.1	Aufgabe	235
7.3.2	Seiten und Sätze	235
7.3.3	Adressierung von Sätzen	237
7.4	<i>Anfrageprozessor</i>	238
7.4.1	Basisoperatoren	238
7.4.1.1	Unäre Operatoren	239
7.4.1.2	Binäre Operatoren	241
7.4.2	Anfrageplanung und -optimierung	242
7.4.3	Kosten und Statistiken	247