

eXamen.press

eXamen.press ist eine Reihe, die Theorie und Praxis aus allen Bereichen der Informatik für die Hochschulausbildung vermittelt.

Andreas Meier

Relationale und postrelationale Datenbanken

6. überarbeitete und erweiterte Auflage
Mit 119 Abbildungen

 Springer

Andreas Meier

Departement für Informatik
Universität Fribourg
Bd. de Perolles 90
1700 Fribourg, Schweiz
andreas.meier@unifr.ch

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.

Bis zur 5. Auflage erschienen im Springer-Verlag Berlin Heidelberg unter dem Titel
Relationale Datenbanken. Leitfaden für die Praxis.

ISSN 1614-5216

ISBN 978-3-540-46553-9 Springer Berlin Heidelberg New York

ISBN 978-3-540-00905-4 5.Auflage Springer Berlin Heidelberg New York

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media
springer.de

© Springer-Verlag Berlin Heidelberg 1991, 1994, 1997, 2001, 2004, 2007

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. Text und Abbildungen wurden mit größter Sorgfalt erarbeitet. Verlag und Autor können jedoch für eventuell verbliebene fehlerhafte Angaben und deren Folgen weder eine juristische Verantwortung noch irgendeine Haftung übernehmen.

Satz: Druckfertige Daten des Autors
Herstellung: LE-TEX, Jelonek, Schmidt & Vöckler GbR, Leipzig
Umschlaggestaltung: KünkelLopka Werbeagentur, Heidelberg
Gedruckt auf säurefreiem Papier 33/3142 YL - 5 4 3 2 1 0

Für Heiri, Ramani und Tina

Geleitwort

Dass ein Fachbuch der Informatik mehrere Auflagen in kurzer Zeit erlebt, ist eine Seltenheit. Die Frage, die zu stellen ist, lautet: Wie kommt diese Ausnahmesituation zustande?

Eine zutreffende, aber leider nur einfach aussehende Antwort wäre: Das Buch von Andreas Meier ist für Praktiker und Lernende verständlich geschrieben, wobei wir verständlich als praktisch nachvollziehbar deuten. Warum, so lautet eine bohrende Zusatzfrage, ist das Buch von Meier nachvollziehbar? Meine Antwort auf diese Frage ist etwas umfangreicher:

Das Buch ist für den Lernenden nachvollziehbar, weil es einsichtig, angemessen und richtig ist. «Einsichtig» soll in diesem Zusammenhang bedeuten, dass das Buch den Anfangswissensstand des Lesers berücksichtigt. Sprache und Bild harmonisieren in dem Buch. Höchste und präzise Abstraktionen sind für einen lernenden Anfänger nicht einsichtig. «Angemessenheit» verlangt, dass eine Darstellung für den Lernenden nach Umfang und Tiefe in den Kontext passt, in dem sie gefordert wird. Akademische Forschung und industrielle Anwendung sind in diesem Sinne zwei grundsätzlich verschiedene Kontexte. Meier hat die industrielle Anwendung im Sinn. «Richtig» ist für jeden Autor der heikelste und anstrengendste Punkt, wenn Richtigkeit nicht einfach als logische Wahrheit, sondern als Übereinstimmung mit der anerkannten Modellbildung verstanden wird, die sich ständig ändert.

In der Datenbankwelt, die aus den alten Dateisystemen der Betriebssysteme heraus geboren wurde, findet seit über dreißig Jahren eine rege Modell-Evolution statt. Über den CODASYL-Ansatz gelangten wir zum Relationen-Modell und stehen nun mit dem Internet vor dem größten vorstellbaren Datenbanksystem. Die Konsequenzen für die Zukunft sind noch nicht absehbar. Gesagt werden kann nur, dass kein ruhiges Fahrwasser vor uns liegt. Es ist deshalb wichtig, dass der Autor in der überarbeiteten Auflage der Integration heterogener Datenbanken im Web Gewicht beimisst und mit einem Kapitel über postrelationale Datenbanken die Zukunftsperspektiven aufzeigt.

Ich wünsche der neuen Auflage eine breite und interessierte Leserschaft.

Erlangen, im November 2006

Hartmut Wedekind

Vorwort zur sechsten Auflage

Wir kennen den sektoralen Strukturwandel, d.h. die längerfristige Verlagerung der Beschäftigung von der Landwirtschaft über den Produktionssektor hin zu Dienstleistungen und zur Informationsverarbeitung. Bereits heute dominieren die Informationsberufe gegenüber anderen Berufssparten deutlich. Immer mehr Personen beschäftigen sich mit der Produktion, Verarbeitung und Verteilung von Informationen. Die westlichen Länder befinden sich auf dem Weg zu einer Wissens- und Informationsgesellschaft.

Datenbanksysteme mit ihren Sprach- und Auswertungsfunktionen zählen neben der Kommunikationsinfrastruktur zu den Schlüsseltechnologien unserer Wirtschaft. Im elektronischen Geschäft benötigen wir webbasierte Produktkataloge für die Informationsphase. Zudem verwenden wir Kundendatenbanken in der Offertstellung, in der Vertragsverhandlung und bei der Kontraktabwicklung. Auch die Distribution der Produkte überwachen wir mit Online-Datenbanken, unabhängig davon, ob es sich um digitale oder materielle Güter handelt. Das kundenindividuelle Marketing ohne Datenbank oder Data Warehouse ist kaum mehr vorstellbar.

Das vorliegende Fachbuch zeigt das gesamte Anwendungsspektrum vom relationalen Datenbankentwurf bis hin zur Entwicklung von postrelationalen Datenbanksystemen im Überblick und praxisnah. Als *fundierte Einführung in das Gebiet der relationalen und postrelationalen Datenbanksysteme* ermöglicht es dem interessierten Praktiker,

- bei der Datenmodellierung und in der Datenbankentwurfsmethodik das Entitäten-Beziehungsmodell und das Relationenmodell anzuwenden,
- relationale Abfrage- und Manipulationssprachen kennenzulernen und einzusetzen,
- die automatischen Abläufe und Techniken innerhalb eines relationalen oder postrelationalen Datenbanksystems zu verstehen,
- Konzepte für die Integration von Datenbanken ins Web und für die Migration zu bewerten sowie

- die Stärken und Schwächen relationaler Technologie zu erkennen, in seine Überlegungen miteinzubeziehen und künftige Entwicklungen einigermaßen richtig abzuschätzen.

Das Fachbuch richtet sich an *Praktiker und Ausbildungsverantwortliche von Unternehmen, an Dozierende und Studierende von Universitäten, Fachhochschulen und Berufsakademien* sowie an alle, die eine praxisbezogene Einführung in die relationale und postrelationale Datenbanktechnik suchen. Im Zentrum stehen wichtige Begriffe, die zwar vielerorts gebraucht, jedoch selten richtig verstanden und angewendet werden. Illustriert wird das Fachbuch durch eine große Anzahl einfacher, meist selbsterklärender Abbildungen. Ein gezielt zusammengestelltes Literaturverzeichnis verweist den interessierten Leser auf weiterführende Publikationen, die einen vertieften Einblick in die angeschnittenen Themenkreise bieten.

Diese Einführung geht die relationale und postrelationale Datenbanktechnologie von verschiedenen Blickwinkeln an. Sie erläutert Entwurfsmethoden und Sprachaspekte ebenso wie wichtige Architekturkonzepte von Datenbanksystemen. Auf eine Beschreibung spezifischer Herstellerprodukte wird verzichtet, um den *grundlegenden Methoden und Techniken mehr Platz einzuräumen* und das Verständnis beim Benutzen relationaler und postrelationaler Datenbanken zu fördern. Somit schließt diese Veröffentlichung eine Lücke in der praxisbezogenen Fachliteratur.

Neu hinzugekommen sind in der sechsten, überarbeiteten und erweiterten Auflage ein Tutorium in SQL (Structured Query Language). Damit wird dem Wunsche vieler Leserinnen und Leser Rechnung getragen, der international standardisierten Sprache mehr Gewicht zu verleihen. Ein Repetitorium sowie ein Appendix mit einer Fallstudie für das Softwaresystem Access erlauben außerdem, die Kenntnisse in SQL konkret zu prüfen. Weitere Übungsaufgaben und Lösungsvorschläge finden sich außerdem auf der neu erstellten Website <http://www.RelationaleDatenbanken.ch>.

Alle Kapitel sind überarbeitet und teilweise erweitert worden. So findet sich neu in Kapitel 1 eine Einführung in das Gebiet der Informationssysteme: Hier wird argumentiert, dass Informationen spezifische Eigenschaften besitzen und als Produktionsfaktor besondere Herausforderungen an Unternehmen und Organisationen stellen. In Kapitel 2 wurde der Teil über die Normalformen vervollständigt, obwohl höhere Normalformen in der Praxis selten anzuwenden sind. Bei den Kapiteln 4 und 5 wurde der Tatsache Rechnung getragen, dass webbasierte Technologien und Lösungen immer mehr Fuß fassen. Schließlich konnte in Kapitel 6 bei der Diskussion von fuzzy Datenbanken ein attraktives Beispiel aus dem Kundenbeziehungsma-

nagement gewählt werden. Damit lassen sich Konflikte bei scharfen Kundensegmenten im Marketing beheben.

Das Fachbuch ist im Rahmen eines firmeninternen Ausbildungsprogramms im Bankenumfeld entstanden, ergänzt durch Diskussionsbeiträge aus den Vorlesungen «Praxis relationaler Datenbanken» an der ETH in Zürich (für Informatiker und Ingenieure) sowie «Informationssysteme und Datenbanken» an der Universität Fribourg (für Studierende der Wirtschaftswissenschaften). Viele Fachkolleginnen und -kollegen aus Praxis und Hochschule haben dazu beigetragen, den Text verständlicher und die Abbildungen anschaulicher zu gestalten. Mein Dank richtet sich an Urs Bebler, Eirik Danielson, Bernardin Denzel, Emmerich Fuchs, Peter Gasche, Caroline Grässle-Mutter, Michael Hofmann, Stefan Hüsemann, Günther Jakobitsch, Hans-Peter Joos, Klaus Küspert, Gitta Marchand, Michael Matousek, Thomas Myrach, Mikael Norlindh, Michel Patcas, Fabio Patocchi, Ernst-Rudolf Patzke, Thomas Rätz, Marco Savini, Werner Schaad, August Scherrer, Walter Schnider, Henrik Stormer, Max Vetter, Hartmut Wedekind und Gerhard Weikum. Ein besonderes Kompliment richte ich an Andreea Ionas, die unermüdlich Text und Zeichnungen redigiert hat. Dem Springer-Verlag, vor allem Clemens Heine, Agnes Herrmann und Michael Reinfarth, danke ich für die angenehme Zusammenarbeit.

Fribourg, im November 2006

Andreas Meier

Inhaltsverzeichnis

1 Der Weg zum Datenmanagement	1
1.1 Der Untersuchungsgegenstand der Wirtschaftsinformatik	1
1.2 Grundbegriffe des Relationenmodells	4
1.3 Die international standardisierte Sprache SQL	7
1.4 Die Komponenten eines relationalen Datenbanksystems	10
1.5 Zur Organisation des Datenbankeinsatzes	12
1.6 Bemerkungen zur Literatur	14
2 Schritte zur Datenmodellierung	17
2.1 Von der Datenanalyse zur Datenbank	17
2.2 Das Entitäten-Beziehungsmodell	20
2.2.1 Entitäten und Beziehungen	20
2.2.2 Assoziationstypen	22
2.2.3 Generalisation und Aggregation	24
2.3 Das relationale Datenbankschema	28
2.3.1 Überführen des Entitäten-Beziehungsmodells	28
2.3.2 Abbildungsregeln für Beziehungsmengen	30
2.3.3 Abbildungsregeln für Generalisation und Aggregation	35
2.4 Abhängigkeiten und Normalformen	38
2.4.1 Sinn und Zweck von Normalformen	38
2.4.2 Funktionale Abhängigkeiten	41
2.4.3 Transitive Abhängigkeiten	43
2.4.4 Mehrwertige Abhängigkeiten	45
2.4.5 Verbundabhängigkeit	48
2.5 Strukturelle Integritätsbedingungen	51
2.6 Unternehmensweite Datenarchitektur	53
2.7 Rezept zum Datenbankentwurf	57
2.8 Bibliografische Angaben	59
3 Abfrage- und Manipulationssprachen	61
3.1 Benutzung einer Datenbank	61
3.2 Grundlagen der Relationenalgebra	63
3.2.1 Zusammenstellung der Operatoren	63
3.2.2 Die mengenorientierten Operatoren	65
3.2.3 Die relationenorientierten Operatoren	68



3.3	Relational vollständige Sprachen	73
3.4	Übersicht über relationale Sprachen	75
3.4.1	SQL	75
3.4.2	QUEL	78
3.4.3	QBE	79
3.5	Eingebettete Sprachen	82
3.6	Behandlung von Nullwerten	83
3.7	Datenschutzaspekte	85
3.8	Formulierung von Integritätsbedingungen	88
3.9	Bibliografische Angaben	91
4	Elemente der Systemarchitektur	93
4.1	Wissenswertes über die Systemarchitektur	93
4.2	Übersetzung und Optimierung	96
4.2.1	Erstellen eines Anfragebaumes	96
4.2.2	Optimierung durch algebraische Umformung	98
4.2.3	Berechnen des Verbundoperators	101
4.3	Mehrbenutzerbetrieb	104
4.3.1	Der Begriff der Transaktion	104
4.3.2	Serialisierbarkeit	106
4.3.3	Pessimistische Verfahren	109
4.3.4	Optimistische Verfahren	113
4.4	Speicher- und Zugriffsstrukturen	115
4.4.1	Baumstrukturen	115
4.4.2	Hash-Verfahren	118
4.4.3	Mehrdimensionale Datenstrukturen	121
4.5	Fehlerbehandlung	124
4.6	Die Systemarchitektur im Detail	126
4.7	Bibliografische Angaben	128
5	Integration und Migration von Datenbanken	131
5.1	Zur Nutzung heterogener Datenbestände	131
5.2	Datenbanken im Web	132
5.2.1	Aufbau eines webbasierten Informationssystems	133
5.2.2	XML-Dokumente und XML-Schemas	134
5.2.3	Die Abfragesprache XQuery	137
5.3	Abbildungsregeln für Integration und Migration	139
5.3.1	Abbildungen für einfache Entitätsmengen und Wiederholungsgruppen	139
5.3.2	Abbildungen für abhängige Entitätsmengen	142
5.3.3	Indirekte Abbildungen für die Datenintegration und -migration	145
5.4	Migrationsvarianten für heterogene Datenbanken	147
5.4.1	Charakterisierung unterschiedlicher Migrationsvarianten	148
5.4.2	Systemkonforme Spiegelung von Datenbanken	150
5.5	Grundsätze der Integrations- und Migrationsplanung	153



5.6	Bibliografische Angaben	156
6	Postrelationale Datenbanksysteme	159
6.1	Weiterentwicklung – weshalb und wohin?	159
6.2	Verteilte Datenbanken	160
6.3	Temporale Datenbanken	165
6.4	Objektrelationale Datenbanken	169
6.5	Multidimensionale Datenbanken	173
6.6	Fuzzy Datenbanken	178
6.7	Wissensbasierte Datenbanken	184
6.8	Literatur zur Forschung und Entwicklung	188
	Repetitorium	191
	Tutorium in SQL	201
	Eine Datenbank mit Access erstellen	217
	Glossar	239
	Fachbegriffe englisch/deutsch	245
	Literaturverzeichnis	249
	Stichwortverzeichnis	259