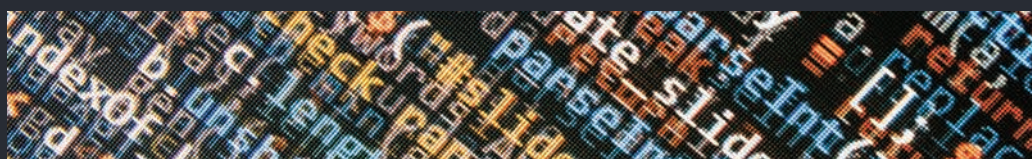
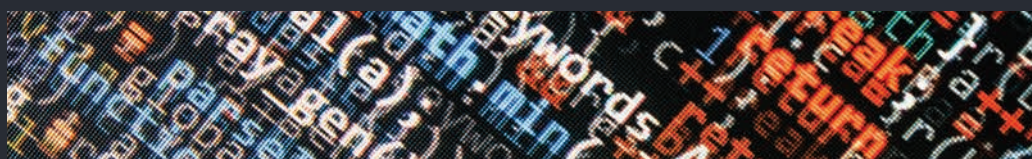
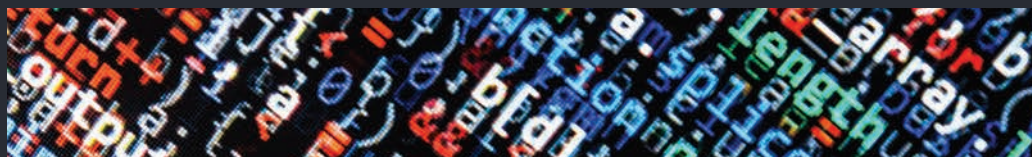


DATABASES EN SQL



5^e DRUK

TON DE ROOIJ

Databases en SQL

Databases en SQL

Vijfde, herziene druk

Ton de Rooij

Boom

Opmaak binnenwerk: T. de Rooij, Woerden
Basisontwerp omslag: Dog e Pony, Amsterdam
Omslagontwerp: Emma Raben, Amsterdam
Beeld omslag: Best-Backgrounds/Shutterstock

© 2023 T. de Rooij | Boom

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bijvoorbeeld een (digitale) leeromgeving of een reader in het onderwijs (op grond van artikel 16, Auteurswet 1912) kan men zich wenden tot Stichting Uitgeversorganisatie voor Onderwijslicenties, Postbus 3060, 2130 KB Hoofddorp, www.stichting-uvo.nl.

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 9789024457182
ISBN e-book 9789024457199
NUR 173

www.boomhogeronderwijs.nl

Voorwoord

Dit boek heeft tot doel dat de lezer:

- inzicht krijgt in database-managementsystemen en inzicht in de taal SQL voor beschrijving en manipulatie van databases
- om kan gaan met de begrippen van het relationele model
- het ontwerp (relationeel schema) van een relationele database kan interpreteren, beoordelen en aanpassen
- SQL kan gebruiken voor het implementeren en manipuleren van een relationele database alsmede het uitvragen ervan.

Het boek richt zich op relationele databases en is geschikt voor zelfstudie, cursorisch onderwijs en als naslagwerk. De inhoud sluit aan bij exameneisen van I-Tracks, waaronder Databases en SQL Foundation (DBSQLF), opgesteld door EXIN, het exameninstituut voor ICT'ers. Deze exameneisen zijn te raadplegen via www.exin.nl.

De eerste helft van het boek behandelt de werking van een DBMS en het relationele model. Er wordt van uitgegaan dat een relationeel ontwerp ontstaat door een informatiemodel om te zetten in een relationeel schema. Normaliseren wordt wel behandeld, maar voor het opstellen van een relationele database wordt niet uitgegaan van normaliseren.

De hoofdstukken 3, 4 en 5 geven een volledige behandeling van het relationele model. Aan de hand hiervan kan van een relationeel schema worden beoordeeld of het voldoet aan:

- de eisen van het relationele model
- de eisen om te komen tot een goed relationeel databaseontwerp.

De tweede helft van het boek richt zich op het opstellen van SQL-opdrachten waarmee relationele databases geraadpleegd en gewijzigd kunnen worden. Daarbij komen uitgebreid strategieën aan de orde die helpen bij het opstellen van queries.

Het boek is modulair opgezet. De hoofdstukken zijn zo georganiseerd dat alles over een onderwerp bij elkaar staat. Daarnaast kunnen delen van het boek los van elkaar bestudeerd worden:

- hoofdstuk 2 over database-managementsysteem
- hoofdstuk 3-5 over het relationele model
- hoofdstuk 7-10 over het opstellen van eenvoudige tot complexe queries.

De vierde druk is verbeterd ten opzichte van eerdere drukken. Er is een duidelijke uitleg toegevoegd over outer joins; daarnaast is de uitleg uitgebreid van de afhankelijkheidsregel, BCNF, en het voorbeeld van de subquery met wederzijdse vervanging van werknemers. Verder zijn een aantal correcties doorgevoerd en zijn enkele voorbeelden aangepast ten behoeve van de leesbaarheid.

De vijfde druk bevat vele aanpassingen. Referenties aan verouderde methoden en uitleg van inmiddels niet meer actuele database-management systemen zijn verwijderd. In de vorige drukken werd zoveel mogelijk gewerkt met Nederlandstalige namen voor relationele begrippen. Het is echter niet meer gebruikelijk om over primaire sleutels, vreemde sleutels en kandidaatsleutels te spreken. Deze namen zijn Engelstalig gemaakt: primary keys, foreign keys en candidate keys. Tot slot is de inhoud van de voorbeelddatabases qua datums aangepast.

Oefenmateriaal

Bij dit boek hoort oefenmateriaal, dat bestaat uit een aantal vragen en een oefendatabase in Access om de vragen op te lossen. Dit oefenmateriaal is als download beschikbaar via de website van de auteur, www.tonderooij.com.

Woerden, mei 2023
Ton de Rooij

Inhoud

Voorwoord		v
1	Inleiding	1
	1.1 Activiteiten informatiesysteemontwikkeling	1
	1.2 Uitleg en aansluiting op dit boek	2
	Samenvatting	5
2	Database-managementsystemen	7
	2.1 Wat zijn databases?	7
	2.2 De werking van een DBMS	8
	2.3 Database-managementsystemen (DBMS'en)	14
	Samenvatting	19
3	Relationeel model: de begrippen	21
	3.1 Algemene begrippen	21
	3.2 Keys / Sleutels	25
	3.3 Integriteit	32
	3.4 Null-waarden	33
	3.5 Consistentie	34
	3.6 Redundantie	34
	Samenvatting	35
4	Relationeel model: ontwerp	37
	4.1 Inleiding	37
	4.2 Ontwerpaspecten	38
	4.3 Normaliseren	44
	Samenvatting	49
5	Relationeel model: verifiëren	51
	5.1 Inleiding	51
	5.2 Aan de hand waarvan verifiëren?	51
	5.3 Wat dwingt het relationele model zelf af?	52
	5.4 Wat moet er verder worden afgedwongen?	54
	Samenvatting	55
6	SQL: implementatie van een relationele database	57
	6.1 Database-definitie	57
	6.2 Voorbeeld-databases	65
	6.3 Grant en Revoke (DCL)	69
	Samenvatting	70

7	SQL: queries met en zonder joins	71
7.1	Gegevensstructuur en programmeren	71
7.2	De SELECT-opdracht en joinen	75
	Samenvatting	91
8	SQL: queries met functies en GROUP BY/HAVING	93
8.1	Functies	93
8.2	GROUP BY-clausule	98
8.3	HAVING-clausule	102
	Samenvatting	105
9	SQL: queries met subqueries	107
9.1	Inleiding	107
9.2	Subquery zonder correlatie	109
9.3	Subquery met correlatie	117
	Samenvatting	126
10	SQL: het gebruik van [NOT] EXISTS	127
10.1	Werking en gebruik van EXISTS en NOT EXISTS	127
10.2	NOT EXISTS gebruikt voor dubbele ontkenningen	131
	Samenvatting	141
11	SQL: update-opdrachten	143
11.1	Inleiding	143
11.2	INSERT	144
11.3	DELETE	145
11.4	UPDATE	146
	Samenvatting	147
	Literatuurlijst	149
	Register	151

Hoofdstuk 1

Inleiding

Dit boek draagt een aantal bouwstenen aan bij het ontwikkelen van een geautomatiseerd informatiesysteem. In dit hoofdstuk geven we aan hoe dit boek bij het ontwikkelen van een geautomatiseerd informatiesysteem kan worden gebruikt. In paragraaf 1.1 geven we aan uit welke activiteiten het ontwikkelen van een geautomatiseerd informatiesysteem bestaat. In paragraaf 1.2 geven we een korte uitleg bij deze activiteiten. Verder geven we in paragraaf 1.2 aan hoe de onderwerpen in dit boek bij de ontwikkelactiviteiten kunnen worden gebruikt.

1.1 Activiteiten informatiesysteemontwikkeling

Het ontwikkelen van een geautomatiseerd informatiesysteem bestaat uit de volgende activiteiten:

- bedrijfsanalyse
- systeemspecificatie
- systeembouw.

Bedrijfsanalyse

Onder bedrijfsanalyse verstaan we het in kaart brengen van het probleemgebied. Hierbij komt aan de orde:

- het afbakenen van het probleemgebied
- het bestuderen van de bedrijfsactiviteiten
- het beschrijven van de informatiebehoeften (vastgelegd in een informatiemodel).

Systeemspecificatie

Onder systeemspecificatie verstaan we het maken van een ontwerp van het te bouwen informatiesysteem. Hierbij komt aan de orde:

- het afbakenen van een informatiesysteem
- het vaststellen welke functies in het informatiesysteem moeten worden opgenomen
- het vaststellen welke gegevens moeten worden vastgelegd (gegevensmodel)
- het vaststellen hoe in de functies van het informatiesysteem zal worden voorzien (software-ontwerp)
- het vaststellen hoe in de opslag van gegevens zal worden voorzien (database-ontwerp).

Systeembouw

Onder systeembouw verstaan we:

- het realiseren van de gegevensopslag (database-definitie)
- het realiseren van de systeemfuncties (programmeren, genereren).

1.2 Uitleg en aansluiting op dit boek

We zullen kort ingaan op ieder van de in paragraaf 1.1 genoemde activiteiten bij het ontwikkelen van een informatiesysteem. We geven bij de uitleg aan hoe de onderwerpen in dit boek op deze activiteiten aansluiten.

Bedrijfsanalyse: afbakenen van het probleemgebied

Het eerste wat aan de orde moet komen is de afbakening van het probleemgebied. De afbakening wordt gevormd door een globale beschrijving van het (deel van het) bedrijfsproces van een organisatie waar we een informatiesysteem voor ontwikkelen.

Bedrijfsanalyse: bestuderen van de bedrijfsactiviteiten

Teneinde een beeld te krijgen van hoe het te ontwikkelen informatiesysteem de activiteiten in een bedrijf gaat ondersteunen, moeten we vaststellen welke bedrijfsactiviteiten (c.q. bedrijfsprocessen) ondersteund moeten gaan worden. Er zijn diverse methoden waarmee we (grafische) modellen van de activiteiten kunnen maken (bijvoorbeeld ACTIMOD [SADT]). De modellen van de activiteiten van een bedrijf gebruiken we voor:

- het beschrijven van de informatiebehoeften (onderdeel bedrijfsanalyse)
- het bepalen van de systeemfuncties die we in een informatiesysteem moeten opnemen (onderdeel systeemspecificatie).

Bedrijfsanalyse: beschrijven van de informatiebehoeften

Bij het uitvoeren van bedrijfsactiviteiten is er behoefte aan informatie. Het bepalen welke informatiebehoeften er zijn (in een bedrijfsproces) is de belangrijkste activiteit bij bedrijfsanalyse.

Er bestaan diverse methoden om te bepalen welke informatiebehoeften er zijn. Met deze methoden stellen we van de informatiebehoeften (grafische) modellen op (voorbeeld van zo'n methode is ERM). Informatiemodellen dan wel datamodellen gebruiken we voor:

- het vaststellen welke gegevens moeten worden vastgelegd (gegevensmodel; zie hierna)
- het vaststellen hoe we in de opslag van gegevens kunnen voorzien (database-ontwerp; een database is een geïntegreerde verzameling gegevens, zie voor een uitgebreide definitie hoofdstuk 2)
- het toetsen op de volledigheid van de werking van een informatiesysteem (voorziet het informatiesysteem wel in alle informatiebehoeften die in het informatiemodel zijn opgenomen?).

Systeemspecificatie: afbakenen van een informatiesysteem

Lang niet alle informatie die we gebruiken in het bedrijfsproces, verwerken we in een geautomatiseerd informatiesysteem. Veel informatie communiceren we mondeling. Veel informatie leggen we niet vast, maar onthouden we. Andere informatie zetten we (ten behoeve van communicatie en onthouden) alleen op

papier. Het is daarom belangrijk af te bakenen wat binnen en wat buiten het geautomatiseerde informatiesysteem valt.

Systeemspecificatie: welke functies in het informatiesysteem?

Een informatiesysteem moet in allerlei functies voorzien. Voorbeelden: het geven van een overzicht van klanten die te laat zijn met betalen, het invoeren van een nieuwe klant, het bewerken van klantgegevens, het opnemen van een nieuwe order. Van de door het informatiesysteem te ondersteunen functies moeten we een overzicht maken.

Een hulpmiddel dat we daarvoor soms gebruiken, is het maken van een (grafisch) model van alle functies (bijvoorbeeld met behulp van DFD's [data flow diagrams] of functiehiërarchieën).

Systeemspecificatie: welke gegevens moeten worden vastgelegd?

Het vaststellen welke gegevens moeten worden vastgelegd, doen we aan de hand van een informatiemodel. Bij iedere informatiemodelleringsmethode is hiervoor een omzettingsprocedure opgenomen. Met zo'n omzettingsprocedure zetten we een model van de informatiebehoeften om in een model van de gegevens waarmee in de informatiebehoeften wordt voorzien. Zie voor het verschil tussen informatie in een informatiemodel en gegevens in een gegevensmodel de voorbeelden in tabel 1.1.

Informatie(-behoefte)	Gegeven (weergave)
hoeveelheid	getal van 10 posities om het aantal weer te geven
hoe heet iemand	een naam van 50 posities
wanneer is iemand geboren	een geboortedatum
bedrag	getal van 10 posities met 2 posities achter de komma

Tabel 1.1 Voorbeeld van informatie en de gegevens die de informatie weergeven

De uitvoer van zo'n omzettingsprocedure is meestal een relationeel schema (een gegevensmodel opgesteld volgens de regels van het relationele model). Aan het relationele schema moeten we allerlei details toevoegen om het geschikt te maken als database-ontwerp (zie hierna bij 'Systeemspecificatie: hoe wordt in de opslag van de gegevens voorzien?').

Wat een relationeel schema inhoudt en hoe we dit in detail ontwerpen, wordt beschreven in de hoofdstukken 3 tot en met 5.

Systeemspecificatie: hoe wordt in de functies van het informatiesysteem voorzien?

Het vaststellen hoe we in de functies van het informatiesysteem voorzien, is het software-ontwerp. Veel van de functies kunnen we met de hulpmiddelen die

tegenwoordig bij databasepakketten leverbaar zijn automatisch genereren. Er blijft altijd een stuk over dat niet gegenereerd kan worden. Dit stuk kunnen we met de geavanceerde hulpmiddelen doorgaans snel ontwikkelen. Soms moeten we echter nog programmeren. Vaak zijn er dan al stukken programma beschikbaar, bijvoorbeeld een module om een factuur te maken.

Systeemspecificatie: hoe wordt in de opslag van de gegevens voorzien?

Het vaststellen hoe we in de opslag van de gegevens voorzien, is het database-ontwerp. Een database-ontwerp bevat alle details om een gegevensmodel in een bepaalde database-omgeving te kunnen implementeren. Details betreffen onder andere:

- aantal posities en gegevenssoort van een gegeven
- regels die moeten gelden, zoals identificaties.

In de hoofdstukken 3 tot en met 5 lichten we toe hoe we een relationele database ontwerpen.

Systeembouw: realiseren van de gegevensopslag

Het realiseren van de gegevensopslag is op diverse manieren mogelijk, variërend van het gebruiken van bestanden en spreadsheets tot databases. In het kader van dit boek beschouwen we alleen databases. In hoofdstuk 6 zullen we laten zien hoe we een relationeel database-ontwerp in een database-definitie omzetten.

Systeembouw: realiseren van de systeemfuncties (programmeren, genereren)

Zoals hiervoor al is gezegd, kunnen veel systeemfuncties (met behulp van geavanceerde hulpmiddelen) automatisch gegenereerd worden of zijn er kant en klare functies als programmamodules beschikbaar. Het genereren gebeurt op basis van een uitgebreid relationeel database-ontwerp, en betreft onder andere:

- schermen voor het invoeren, bewerken en opvragen van gegevens
- standaardrapporten, zoals een lijst van alle klanten.

In een aantal van de overblijvende functies voorzien we door gebruik te maken van rapportgeneratoren. Met deze rapportgeneratoren kunnen we allerlei regelmatig terugkomende rapportages ontwerpen en maken.

Weer andere functies kunnen we alleen ontwikkelen door zelf programma's te schrijven. Het betreft dan ingewikkelde vragen. Veel van deze vragen zijn ad hoc-vragen (vragen die plotseling bedacht worden en die we bij het bouwen van het informatiesysteem nog niet kunnen voorzien). Deze vragen liggen meestal op het niveau van de management-informatie (bijvoorbeeld: welk percentage van onze klanten heeft de vorige maand een bestelling gedaan?). De meeste informatie die voor het operationele bedrijfsproces nodig is (bijvoorbeeld welke artikelen moeten worden ingepakt om order 14361 uit te leveren), kan verstrekt worden met de schermen die door de hiervoor genoemde hulpmiddelen zijn gemaakt.

In hoofdstuk 6 tot en met 11 gaan we diepgaand in op het schrijven van programma's voor een relationele database. Hierbij gebruiken we de programmeertaal SQL.

Samenvatting

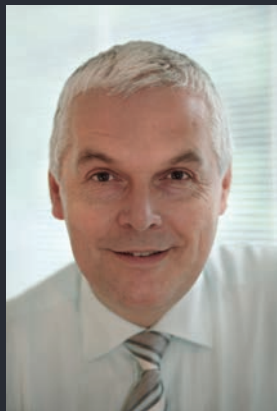
In dit hoofdstuk is aan de orde gekomen hoe de inhoud van dit boek aansluit op onderdelen van het ontwikkeltraject voor het bouwen van informatiesystemen. We hebben gezien uit welke activiteiten en deelactiviteiten het bouwen van informatiesystemen bestaat. We hebben daarbij gezien dat het boek aansluit op deelactiviteiten van systeemspecificatie en systeembouw.

Het hele boek gaat over het werken met databases. Voor een goed begrip van het werken met databases moet u weten wat een database is. In hoofdstuk 2 gaan we hierop uitgebreid in. Bovendien gaan we in dit hoofdstuk in op de software-omgevingen (database-managementsystemen) om databases mee op te zetten en te onderhouden.

In hoofdstuk 3 tot en met 5 leggen we uit wat het relationele model inhoudt en hoe we met behulp van het relationele model een relationeel database-ontwerp kunnen maken. Hoofdstuk 6 bevat uitleg over hoe we een relationele database met behulp van SQL-opdrachten kunnen implementeren. De hoofdstukken 7 tot en met 10 beschrijven diepgaand hoe we informatie uit een relationele database met behulp van de taal SQL kunnen ontsluiten. Hoofdstuk 11 laat zien hoe we de inhoud van een relationele database kunnen wijzigen.

Databases en SQL behandelt de plaats van databasetechnologie in systeemontwikkeling, wat databasemanagementsystemen zijn, en het relationele model waar relationele databases op gebaseerd zijn. Achtereenvolgens is er aandacht voor het relationele begrippenkader, het ontwerpen van een relationele database en de verificatie van het ontwerp. Ook wordt getoond hoe met normaliseren de structuur bepaald kan worden. Verder behandelt het boek de implementatie van een relationele database met behulp van SQL.

Een belangrijk onderdeel van het boek is het (leren) maken van queries. Er zijn aparte hoofdstukken die ingaan op queries waarin al dan niet gebruikgemaakt wordt van joins en het werken met GROUP BY/HAVING. Het onderwerp subqueries is verdeeld over drie hoofdstukken: een hoofdstuk zonder correlatie, een hoofdstuk met correlatie en een hoofdstuk met complexe queries waarin [NOT] EXISTS gebruikt wordt.



De vijfde druk is geactualiseerd en met name het deel over datamanagementsystemen is grondig herzien.

Databases en SQL is voor studenten van opleidingen informatica in het hoger onderwijs en voor bijscholing van professionals.

Ton de Rooij geeft cursussen op het gebied van het ontwikkelen van informatiesystemen: bedrijfsanalyse, datamodellering, databases, SQL en datawarehouses. Daarnaast doet hij advieswerk en maakt hij analyses en ontwerpen voor informatiesystemen.

