

BEN GROENENDIJK

Boom

Aan de slag met Power BI

VAN DATA NAAR DASHBOARD

DERDE DRUK



Aan de slag met Power BI

Aan de slag met Power BI

Van data naar dashboard

Derde druk

Ben Groenendijk

Boom

+ Online leeromgeving

Met onderstaande unieke activeringscode krijg je via www.boomstudent.nl toegang tot de online leeromgeving met extra materiaal. Deze code is persoonsgebonden en gekoppeld aan de derde druk. Na activering van de code is de online leeromgeving twee jaar toegankelijk. De code kan tot zes maanden na het verschijnen van een volgende druk geactiveerd worden. De code is eenmalig te gebruiken.

Opmaak binnenwerk: Holland Graphics, Amsterdam

Basisontwerp omslag: Dog & Pony, Amsterdam

Omslagontwerp: Coco Bookmedia, Amersfoort

Beeld omslag: KanawatTH (Adobe Stock)

© Ben Groenendijk & Boom uitgevers Amsterdam, 2023

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden vervaelvoudigd, 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 zover het maken van reprografische vervaelvoudingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.stichting-pro.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 9789024462698

ISBN 9789024462704 (e-book)

NUR 123

www.boomstudent.nl

www.boomhogeronderwijs.nl

Voorwoord

Informatie over Power BI is zeer gefragmenteerd te vinden op internet. Boeken over Power BI zijn erg gedetailleerd en gaan in op alle mogelijkheden die het programma kent. Voor het leren werken met Power BI zijn beide ongeschikt, er is geen logische opbouw. Je kunt Power BI alleen leren door zelf aan de slag te gaan met de app, hands-on. Dit boek behandelt alleen de essentiële onderwerpen die voor de (toekomstige) professional van belang zijn. Naast een inleidend hoofdstuk waarin de belangrijkste terminologie binnen datawarehousing wordt toegelicht, is het een hands-on boek.

De lezer gaat stap voor stap zelf aan de slag met Power BI, inclusief video-ondersteuning bij alle onderwerpen. Het is geschikt voor diegene die Power BI wil gaan toepassen voor studie of werk. Diegenen die Power BI al gebruiken leren het beter te doorgronden en professionele rapporten te bouwen.

Het is niet noodzakelijk al voorkennis te hebben over databases. In het eerste inleidende hoofdstuk worden begrippen zoals databases, relaties, datamodel, data-analytics, datawarehouses, ETL-proces, kubussen, big data, etc. toegelicht. Niet alleen het bouwen van visuals, presentatie van data, wordt toegelicht, maar ook het aanpassen en laden van data in het datawarehouse. Kunstmatige intelligentie, de DAX-taal en het maken van dashboards worden uitgebreid behandeld.

Alle onderwerpen worden met concrete voorbeelden stap voor stap toegelicht, zijn voorzien van relevante schermafdrucken en video-ondersteuning. De video-ondersteuning is een noodzakelijk onderdeel van het boek (blended-learning). Bij het boek worden vele bestanden geleverd om te oefenen met Power BI. Sommige met miljoenen rijen aan data, zodat de kracht van Power BI getoond kan worden. De bestanden en video's zijn alleen beschikbaar via de website www.boomstudent.nl met een unieke code die in dit boek is afgedrukt. Op die website zijn ook de uitwerkingen van de voorbeelden te downloaden.

Het boek is geschikt voor de Nederlandstalige of Engelstalige versie van Power BI Desktop. In het boek is uitgegaan van de Nederlandstalige versie, maar steeds is ook het Engelstalige equivalent erbij gegeven.

Na het lezen en uitvoeren van de voorbeelden beschik je over alle kennis en tools binnen Power BI die noodzakelijk zijn om professionele rapporten en dashboards te bouwen.

Vragen, opmerkingen of aanvragen voor bedrijfstrainingen en consultancy naar aanleiding van dit boek zijn welkom en kunnen worden gestuurd aan: ben.groenendijk@outlook.com.

Augustus 2023
Ben Groenendijk

Inhoud

	Voorwoord	5
	Inleiding	9
	Opbouw van het boek	9
	Online leeromgeving www.boomstudent.nl	10
1	Data modeling	11
	1.1 Business Intelligence	11
	1.2 Data, de bron van BI	11
	1.3 Data Analytics	12
	1.4 Databases en tabellen	13
	1.5 Relaties tussen tabellen in een database	16
	1.6 Datawarehouse	18
	1.7 Kubussen	20
	1.8 Tabellen en relaties in een datawarehouse	23
	1.9 Big Data	28
2	Power BI en visuals	33
	2.1 Power BI	33
	2.2 Visuals	35
	2.2.1 <i>Visuals met één tabel</i>	35
	2.2.2 <i>Visuals met meerdere tabellen</i>	45
	2.2.3 <i>Visuals met berekende kolommen en metingen</i>	49
	2.3 Publiceren	75
3	Kunstmatige intelligentie	79
	3.1 Uitgevouwen structuur	79
	3.2 Belangrijkste beïnvloeders	86
	3.3 Analyseren	100
4	Extract, Transform en Load proces	105
5	Kardinaliteit en filterrichting	121
	5.1 Een-op-een relaties	121
	5.2 Een-op-veel relaties	126

5.3	Kruisfilterrichting bij relaties	130
5.4	Veel-op-veel relaties	137
5.4.1	<i>Een-veel-veel-een relaties</i>	137
5.4.2	<i>Veel-een-een-veel relaties</i>	138
6	De meest gebruikte DAX-functies	143
6.1	De dimensietabel Datum	143
6.2	Filteren met DAX	154
6.3	Berekeningen in DAX met de functie <i>CALCULATE()</i>	157
6.3.1	<i>CALCULATE() en annuleren filtercontext</i>	158
6.3.2	<i>CALCULATE() en wijzigen filtercontext</i>	161
6.3.3	<i>CALCULATE() en creëren filtercontext</i>	167
6.3.4	<i>CALCULATE() en rijcontext</i>	169
6.4	Tijdintelligentie in DAX	170
6.5	Klassenindeling in DAX	184
6.6	Rangorde in DAX	188
6.7	Variabelen, cumulatieve waarden en Pareto	193
6.8	Vensterfuncties	200
7	Tips, trucs en nuttige visuals	213
7.1	Metingen in één tabel plaatsen	213
7.2	Snelle metingen	215
7.3	Tijdintelligentie met toekomstige datums	217
7.4	Metervisuals	220
7.5	Infographic Designer	223
8	Professionaliseren dashboard en rapport (Reporting)	227
8.1	Bladwijzers	228
8.2	Tooltips met detailinformatie	231
8.3	Drillthrough	235
8.4	Knoppen en stuurtabellen	238
8.5	Dynamische tekst en titels	246
8.6	LeMair-dashboard	252
	Index	253

Inleiding

Data wordt steeds belangrijker. Organisaties krijgen steeds meer data uit verschillende bronnen: data uit databases binnen de eigen organisatie, data uit Excel, mail, Facebook, Instagram, etc. Met Power BI kunnen die data ingelezen worden in een datawarehouse, bewerkt en grafisch getoond worden in rapporten en dashboards. Met name het visualiseren van data (visuals) is belangrijk geworden. Analyses kunnen schijnbaar automatisch via kunstmatige intelligentie (artificial intelligence – AI) uitgevoerd worden binnen AI-visuals. De app Power BI is wereldwijd marktleider om dat te realiseren. Microsoft biedt de app Power BI Desktop gratis aan. Via het web, telefoon of tablet kunnen de rapporten en dashboards getoond worden nadat ze zijn gepubliceerd.

Boeken over Power BI zijn zeer uitputtend en gaan uitgebreid in op alle mogelijkheden van het programma. Op internet zijn vele filmpjes te vinden over het werken met Power BI, maar dit materiaal is zeer gefragmenteerd. Voor het leren werken met Power BI zijn beide ongeschikt, er is geen logische opbouw. Je kunt Power BI alleen leren door zelf aan de slag te gaan met de app, hands-on. In dit boek is de stof met een logische opbouw beschreven, waarbij je zelf de voorbeelden stap voor stap moet uitvoeren in Power BI. Hierdoor krijg je de gewenste vaardigheid. Bij ieder voorbeeld is bovendien een video-instructie voor een toelichting.

Dit boek behandelt de essentiële onderwerpen van Power BI die voor de (toekomstige) professional van belang zijn.

Opbouw van het boek

In dit boek wordt in hoofdstuk 1 een theoretische toelichting gegeven over begrippen binnen datawarehousing. Databases, relaties, datamodel, data-analytics, datawarehouses, ETL-proces, kubussen, big data, etc. worden toegelicht. Zonder deze basiskennis kan geen professionele datawarehouse ingericht worden.

In hoofdstuk 2 ga je diverse visuals in Power BI maken. De benodigde bestanden en video-ondersteuning worden bij het boek geleverd via de online leeromgeving www.boomstudent.nl.

In hoofdstuk 3 worden enkele speciale visuals toegelicht waarmee op basis van kunstmatige intelligentie krachtige analyses gemaakt kunnen worden op de data.



Even belangrijk als het maken van visuals is consistente data in het warehouse van Power BI laden. Zonder juiste data, geen visuals of visuals met onjuiste informatie. Het bewerken van data en vervolgens inlezen in het datawarehouse van Power BI zijn belangrijke handelingen. Dit wordt het ETL-proces (Extract-Transform-Load) genoemd. Hoofdstuk 4 gaat over het ETL-proces. Een datawarehouse beschikt over data uit verschillende bronnen. Naast de databases van de organisatie, data uit Excel, e-mail, Facebook, Instagram, etc. Vaak moet die data bewerkt worden (transform, data cleansing) voordat die geschikt zijn voor het datawarehouse van Power BI.

Binnen een datawarehouse komen alle soorten relaties voor, een-op-een, een-op-veel en veel-op-veel. Bovendien heeft Power BI een filterrichting. Iedere relatievorm heeft zijn bijzondere eigenschappen. Ook de filterrichting kan voor verrassingen zorgen. Kennis hierover is belangrijk. In hoofdstuk 5 worden deze begrippen met concrete voorbeelden toegelicht.

In hoofdstuk 6 wordt dieper ingegaan op de DAX-taal van Power BI en de datumtabel met tijdsintelligentie. De belangrijkste DAX-functies worden uitgebreid toegelicht. Een zeer belangrijk hoofdstuk.

In hoofdstuk 7 worden enkele handige tips uitgewerkt en nog enkele nuttige visuals.

Als laatste draait hoofdstuk 8 om het bouwen van professionele rapporten en dashboards met behulp van bladwijzers, fraaie tooltips, drillthrough opties, knoppen & stuurtabelen en dynamische titels.

Online leeromgeving www.boomstudent.nl



Bij het boek hoort een online leeromgeving. Gebruik de code op de colofonpagina van je boek om in te loggen. Hierop zijn alle bestanden die in het boek gebruikt worden te downloaden. Er zijn bestanden met miljoenen rijen aan data zodat de kracht van Power BI goed getoond kan worden. Naast deze bestanden zijn op het portaal ook de ondersteunde video's beschikbaar bij de verschillende onderwerpen. Tot slot vind je er de uitwerkingen van de opdrachten.

Data modeling

1

Met Power BI kunnen gegevens worden gevisualiseerd. Die gegevens komen uit gemodelleerde tabellen die onderlinge relaties hebben. In dit hoofdstuk wordt de theorie uitgelegd van hoe die gegevenstabellen werken in een datamodel. Hierbij worden veelvoorkomende begrippen en termen die gebruikt worden bij databases en datawarehousing met voorbeelden toegelicht. Het valt buiten het bestek van dit boek om hier diep op in te gaan, maar om goed met Power BI te kunnen werken is kennis van de werking van een datawarehouse en het modeleren van data noodzakelijk. Ook worden in dit hoofdstuk Big Data, Business Intelligence en Data Analytics toegelicht.

Bekijk de video *Introductie Power BI* via www.boomstudent.nl.



1.1 Business Intelligence

Voor het begrip Business Intelligence (BI) zijn vele definities. Als we deze definities samenvatten dan verstaan we onder BI: uit gegevens (data) management-informatie creëren die leidt tot meer kennis en inzicht om betere beslissingen te kunnen nemen. BI wordt gezien als de opvolger van Management Informatie Systemen (MIS). Bij die systemen worden periodiek managementrapportages gemaakt. BI is het proces van het verzamelen van data uit diverse bronnen en hiermee snel de gewenste informatie op een visuele manier tonen. Met visueel worden dashboards en grafische weergaven van informatie bedoeld.

1.2 Data, de bron van BI

Organisaties verzamelen en raadplegen steeds meer data. Data die een organisatie nodig heeft om te functioneren, zoals het registreren van patiëntgegevens, urenregistratie, inkoopgegevens, verkoopgegevens, financiële gegevens, studentgegevens, vastgoedgegevens of gegevens over verkeersstromen.

Die data kunnen worden omgezet in informatie om inzicht te krijgen in de organisatie. Welk verkoopkanaal genereert de meeste omzet? Wat zijn de kosten per afdeling in een zorginstelling? Wat is per tijdsinterval de gemiddelde snelheid op een snelweg? Wat is per ruimte in een gebouw de gemiddelde temperatuur tijdens kantooruren, avonduren en weekends?

Met informatie kan een organisatie ook inzicht krijgen in hoe ze presteert. Op basis van de juiste informatie kan bijvoorbeeld gemonitord worden of je op de goede weg bent op basis van Key Performance Indicators (KPI's). Zijn de doelen bereikt? Dit kunnen kortetermijndoelen zijn, zoals maandcijfers, aantal bezoekers per dag op de Facebookpagina, gemiddeld aantal clicks op een webpagina, of langetermijndoelen, zoals het vergroten van het marktaandeel.

Met informatie kan ook onderzoek gedaan worden om een doelgroep beter te begrijpen. Welk marketingkanaal heeft het meest bijgedragen aan de omzet? Via welk mediakanaal komen de meeste klachten binnen? Hoeveel procent van de orders komt binnen via mobiele telefoon, tablet of laptop?

De data kunnen opgeslagen liggen in databases op lokale computersystemen (on-premise), maar data worden steeds meer opgeslagen in de Cloud. Naast de (basis)data uit informatiesystemen zoals Customer Relationship Management (CRM) of Enterprise Resource Planning (ERP) kunnen de gebruikte data afkomstig zijn uit mediabronnen, zoals e-mail, website, Facebook, Twitter, Instagram, TikTok. Wellicht wordt ook nog gebruikgemaakt van Big Data. Naast de eigen bronnen kunnen via internet beschikbare externe gegevensbronnen nodig zijn, zoals juridische databanken, data via portals van andere organisaties, of landingstijden van de vliegtuigen op Schiphol. Als je al die gegevens uit de verschillende bronnen samenbrengt, ontstaat een datawarehouse.

1.3 Data Analytics

Om aan de gewenste informatie te komen, kan de data uit de aanwezige bronnen geanalyseerd worden (Data Analytics) en getoond worden in dashboards of grafische rapporten. De persoon die zich daarmee bezighoudt, wordt de data-analist genoemd. Power BI is een tool om data analytics mee uit te voeren. Je kunt hiermee uit verschillende databronnen de gewenste managementinformatie visueel tonen via dashboards of grafische rapporten, zie figuur 1.1. Er zijn meerdere tools voor Data Analytics, zoals Tableau, maar volgens onderzoeksbureau Gartner is Power BI de onbetwiste marktleider.



Figuur 1.1

1.4 Databases en tabellen

De meeste gegevens komen uit relationele databases die lokaal of in de Cloud zijn opgeslagen. Dit zijn databases waarin de gegevens van de dagelijkse processen worden opgeslagen. Voor het werken met Power BI is het daarom van belang dat je kennis hebt over hoe gegevens in databases zijn opgeslagen en de werking daarvan. Zonder die kennis begeef je je op glad ijs in Power BI. Het foutief koppelen van gegevens uit databases, datawarehouses en andere bronnen, zoals gegevens uit een Excelbestand, gegevens van een website of gegevens uit e-mail, kan leiden tot onjuiste informatie.

Aan de hand van een voorbeeld zal duidelijk gemaakt worden waarom gegevens in een database worden opgeslagen. Stel, we gaan een lang weekend weg en reserveren een hotel bij Booking.com. Op de site van Booking.com worden incheckdatum, uitcheckdatum, land, stad en een geschikt hotel geselecteerd. Vervolgens een kamertype, aantal kamers, aantal personen per kamer, de persoonsgegevens, e-mailadres en uiteindelijk de betaalgegevens. Na een geslaagde boeking krijg je direct een e-mail met de boekingsgegevens in PDF-formaat. Je zou misschien denken dat Booking.com iedere boeking als PDF-bestand in zijn boekingsstelsel opslaat. De bestandsnaam zou dan het unieke boekingsnummer kunnen zijn, zodat het ook snel teruggevonden kan worden. Het grootste

nadeel van die methode is dat er geen managementinformatie of marketinginformatie uit de PDF-bestanden te halen is. Bijvoorbeeld: wat is de omzet over de afgelopen maand? Wat is per land het aantal boekingen over het afgelopen jaar? Welke hotels hebben het meeste opgebracht in het laatste kwartaal? Op basis van het boekingsgedrag kan Booking.com de klant verleiden met een gerichte e-mail. Bij tien boekingen per maand zouden de benodigde gegevens nog achterhaald kunnen worden door de PDF-bestanden een voor een na te pluizen, maar bij duizenden boekingen per dag is dat onuitvoerbaar! Met een database kan dat wel.

In een database worden gegevens (data) in logische groepen opgeslagen. Een vereenvoudigd databasemodel (datamodel) van ons voorbeeld is weergegeven in figuur 1.2.



Figuur 1.2

Er is een bestand *Klant* met alleen (geregistreerde) klantgegevens. Verder een bestand *Boeking* met alleen boekingsgegevens en een bestand *Hotel* met gegevens van de te boeken hotels. In het bestand *Klant* worden de basisgegevens van alle klanten gestructureerd opgeslagen in de vorm van een tabel, zie figuur 1.3. In databaseterminologie wordt gesproken van tabel *Klant*.

KLANT

Klantnr	Naam	Adres	Postcode	Plaats	E-mail
100001	T. de Vries	Steenstraat 41	1380 VB	Weesp	vries34@gmail.com
100002	D. Aanraad	Leiweg 12	1621 AP	Hoorn	d.aanraad@outlook.com
100003	M. den Hoed	Poststraat 111	3011 VD	Rotterdam	m.den.hoed@bijkorf.nl
.....					

Figuur 1.3

Een tabel heeft rijen en kolommen. In de kolomkoppen van de tabel staan alle gewenste gegevens van een klant. In databases worden die kolomkoppen *velden* genoemd. Zo is er het veld *Klantnr*, het veld *Naam*, enzovoort, tot en met het veld *E-mail*. In een rij staan de klantgegevens van één klant. In een kolom staan alle waarden van het betreffende veld. Zo bevat de kolom *Naam* alle namen van onze klanten. In een relationele database moeten alle rijen uniek zijn. Twee of meer rijen met exact dezelfde informatie opnemen is niet toegestaan. Het veld *Klantnr* zorgt voor unieke rijen in tabel *Klant*. Hierdoor ontstaat, indien ook de

geboortedatum opgenomen zou zijn, ook bij tweelingen die op hetzelfde adres wonen en een gezamenlijk e-mailadres hebben een unieke rij. Het veld dat een rij in een tabel uniek maakt wordt *sleutelveld* genoemd, kortweg sleutel (Primary Key – PK). Om die reden hebben we een patiëntennummer, een studentnummer, een burgerservicenummer, een cliëntnummer of een zaaknummer.

In tabel *Boeking* worden de basisgegevens van een boeking gestructureerd opgeslagen, zie figuur 1.4.

BOEKING

Boekingnr	Hotelnr	Checkin	Checkout	Kamertype	Aant. pers.	Aantal kamers	Commissie	Klantnr
931321100	678891	8-7-2023	10-7-2023	Suite King	2	1	\$ 45,30	100003
931321101	187612	31-12-2023	8-1-2024	DZ standaard	2	1	\$ 213,80	564162
931321102	476324	17-9-2023	19-9-2023	Luxe Twin	1	2	\$ 48,70	761321
.....								

Figuur 1.4

Bij iedere boeking worden de velden *Boekingnr*, *Hotelnr* tot en met *Klantnr* opgeslagen. In het veld *Commissie* zit het bedrag dat Booking.com voor de boeking ontvangt van het hotel. Het veld *Boekingnr* is in tabel *Boeking* de sleutel. Voor de overzichtelijkheid zijn niet alle kenmerken getoond, zoals boekingsdatum, betaaldatum, betaalwijze, enzovoort. Van het geboekte hotel is alleen het hotelnummer opgenomen. Van de klant die geboekt heeft is alleen het klantnummer opgenomen.

In tabel *Hotel* worden de basisgegevens van alle hotels gestructureerd opgeslagen, zie figuur 1.5. Het veld *Hotelnr* is in tabel *Hotel* de sleutel.

HOTEL

Hotelnr	Naam	Adres	Postcode	Plaats	Land
187612	Gletscherblick	Poststrasse 8	6561	Ischgl	Oostenrijk
187613	Rosewood	Verdun 1	7690	Franschhoek	Zuid-Afrika
187614	Ibis Schiphol	Schipholweg 181	1171 PK	Badhoevedorp	Nederland
.....					

Figuur 1.5

De gegevens in de tabellen zijn zo efficiënt mogelijk opgeslagen. Dit wordt een genormaliseerde structuur genoemd, het normaliseren van de gegevens. Zo wordt bij een boeking in tabel *Boeking* alleen het hotelnummer opgeslagen en

niet alle hotelgegevens. Zou je alle hotelgegevens wel iedere keer opslaan in tabel *Boeking*, dan zouden dezelfde hotelgegevens meerdere keren in tabel *Boeking* voorkomen. Als het hotel bijvoorbeeld tien keer wordt geboekt, zouden de hotelgegevens tien keer voorkomen. Dezelfde gegevens worden dan meerdere keren opgeslagen. Dit wordt *redundantie* genoemd. Redundantie is een bekend verschijnsel bij relationele databases, en moet zoveel mogelijk voorkomen worden. Als je dezelfde gegevens meerdere keren opslaat, bijvoorbeeld klantgegevens, en de klant verhuist of krijgt een ander telefoonnummer, moet je die wijzigingen ook op meerdere plaatsen doorvoeren. Als je dat niet doet, ontstaat er een ernstig probleem dat *inconsistentie* wordt genoemd. Inconsistentie betekent tegenstrijdigheid. Er zijn dan bijvoorbeeld verschillende adressen van een klant bekend, waarmee de database onbetrouwbaar wordt. Je denkt misschien dat zoiets niet voorkomt, maar het komt in organisaties veel voor. Bij fusies van organisaties heb je al twee databases en die moeten wel op een juiste manier samengevoegd worden. Vaak lukt dat niet goed. Zelf zie ik bij een niet nader te noemen bank online, bij persoonlijke gegevens, een adres staan waar ik al meer dan tien jaar niet meer woon (adreswijziging doorgegeven). Post van diezelfde bank krijg ik wel op het goede adres. De adresgegevens zijn dus minimaal op twee plaatsen opgeslagen.

1.5 Relaties tussen tabellen in een database

Een relationele database heeft tabellen met gestructureerde gegevens. In een relationele database van een ziekenhuis zitten alle patiëntgegevens in een tabel, alle behandelingen in een tabel, alle medicijnen in een tabel, alle leveranciers in een tabel, enzovoort. In ons (vereenvoudigde) voorbeeld over het boeken van een hotel zitten de gegevens van alle boekingen in drie tabellen in de database. Tussen die tabellen zijn relaties (koppelingen). Tabel *Klant* en tabel *Boeking* zijn gekoppeld via het veld *Klantnr*, zie figuur 1.3 en 1.4. In tabel *Klant* is het veld *Klantnr* de sleutel, in tabel *Boeking* wordt het veld *Klantnr* een *vreemde sleutel* (Foreign Key – FK) genoemd. Het veld *Klantnr* is in tabel *Boeking* niet de sleutel (dat is het veld *Boekingnr*), maar het wijst naar een sleutelveld in een andere (vreemde) tabel, in dit voorbeeld tabel *Klant*. Tabel *Boeking* heeft nog een vreemde sleutel, namelijk het veld *Hotelnr*. Het veld *Hotelnr* in tabel *Boeking* wijst naar een sleutelveld in de (vreemde) tabel *Hotel*. Door de koppeling tussen sleutel en vreemde sleutel ontstaat de relatie tussen tabellen.

Selecteer je een willekeurige rij in tabel *Boeking*, dan wordt automatisch op basis van het klantnummer (sleutel – vreemde sleutel) in de rij van tabel *Boeking* de bijbehorende klant geselecteerd in tabel *Klant*, zie figuur 1.6.

BOEKing

Boekingnr	Hotelnr	Checkin	Checkout	Kamertype	Aant. pers.	Aant. kamers	Commissie	Klantnr
931321100	678891	8-7-2023	10-7-2023	Suite King	2	1	\$ 45,30	100003

KLANT

Klantnr	Naam	Adres	Postcode	Plaats	E-mail
100003	M. den Hoed	Poststraat 111	3011 VD	Rotterdam	m.den.hoed@bijkorf.nl

Figuur 1.6

Maar het werkt ook andersom. Selecteer je een willekeurige rij in tabel *Klant*, dan wordt automatisch op basis van het klantnummer in de rij van tabel *Klant* de bijbehorende boeking(en) geselecteerd in tabel *Boeking*, zie figuur 1.7.

KLANT

Klantnr	Naam	Adres	Postcode	Plaats	E-mail
100003	M. den Hoed	Poststraat 111	3011 VD	Rotterdam	m.den.hoed@bijkorf.nl

BOEKing

Boekingnr	Hotelnr	Checkin	Checkout	Kamertype	Aant. pers.	Aant. kamers	Commissie	Klantnr
927632251	515453	3-2-2021	17-2-2021	Royal Twin	2	1	\$ 264,90	100003
931321100	678891	8-7-2023	10-7-2023	Suite King	2	1	\$ 45,30	100003
931334241	231765	28-12-2023	2-1-2024	Exel Prime	2	1	\$ 132,00	100003

Figuur 1.7

Belangrijk is uiteraard dat de klantnummers in tabel *Boeking* ook bestaan in tabel *Klant*. Anders wordt gerefereerd aan een klant die niet bestaat. Dat principe wordt referentiële integriteit genoemd. De relatie tussen tabel *Klant* en tabel *Boeking* is een-op-veel. Een klant kan meerdere (veel) boekingen hebben, maar een boeking is gekoppeld aan exact één klant. Dit geldt voor alle relaties in een relationele database. Veel-op-veel of een-op-veel komen in een relationele database niet voor. De werking van de relaties tussen de tabellen in een database kun je je voorstellen door de tabellen te vertalen naar Microsoft Excel. In het eerste tabblad in Excel maak je tabel *Klant*, in het tweede tabblad tabel *Boeking* en in het derde tabblad tabel *Hotel*. Alle tabbladen zijn gevuld met tienduizenden regels. Als Excel hetzelfde zou werken als een database en je een willekeurige klant zou selecteren in het tabblad *Klant* en vervolgens het tabblad *Boeking* zou selecteren, dan zou je alleen nog maar de boekingen zien van de geselecteerde

klant uit het tabblad *Klant*. De tienduizenden boekingsrijen zijn dan automatisch gefilterd op het geselecteerde klantnummer uit tabel *Klant*. Stel, er zijn drie boekingen van de geselecteerde klant zichtbaar in het tabblad *Boeking*. Selecteer een van de drie boekingen en selecteer vervolgens het tabblad *Hotel*. Je ziet nu nog maar één regel, namelijk het hotel dat geboekt is. De koppeling (relatie) tussen tabel *Boeking* en tabel *Hotel* loopt via het veld *Hotelnr.* Er wordt in het tabblad *Hotel* van Excel dan automatisch gefilterd op het geselecteerde hotelnummer uit het tabblad *Boeking*.

De relaties tussen de tabellen zijn automatisch wederkerig. Als je bijvoorbeeld een willekeurig hotel zou selecteren in het tabblad *Hotel* en je vervolgens het tabblad *Boeking* selecteert, dan zie je alle boekingen die gedaan zijn voor het hotel dat op het tabblad *Hotel* is geselecteerd. Selecteer je daarna een willekeurige boeking en kijk je vervolgens op het tabblad *Klant*, dan zie je maar één klant. Het is de klant van de geselecteerde boeking uit het tabblad *Boeking*.

Door de gegevensgroepen voor te stellen als tabellen in Excel is de werking van een database goed voor te stellen. Maar uiteraard kan een database niet gerealiseerd worden in Excel, en bovendien is Excel beperkt. Het maximum aantal rijen in Excel is 1.048.576. Booking.com heeft wereldwijd al meer dan 27 miljoen hotels/appartementen/kamers die geboekt kunnen worden, en het aantal boekingen wereldwijd en dus rijen in de gegevensgroep *Boeking* is meer dan 700.000 per dag! Dus die gegevens verwerken gaat in Excel al niet meer lukken. Een database heeft die beperking niet. De enige beperkingen zijn de beschikbare hoeveelheid geheugen en systeembronnen. Nog een belangrijk verschil tussen Excel en een database is dat in een database alle rijen in een tabel uniek zijn! In Excel zou je twee precies dezelfde rijen kunnen invoeren. Maar in een database kan dat niet, iedere rij in de tabel moet uniek zijn. Wordt een unieke regel opnieuw ingevoerd, dan geeft een database automatisch een foutmelding. De rijen in een database worden uniek gemaakt door het toekennen van unieke klantnummers, boekingsnummers, hotelnummers, enzovoort.

1.6 Datawarehouse

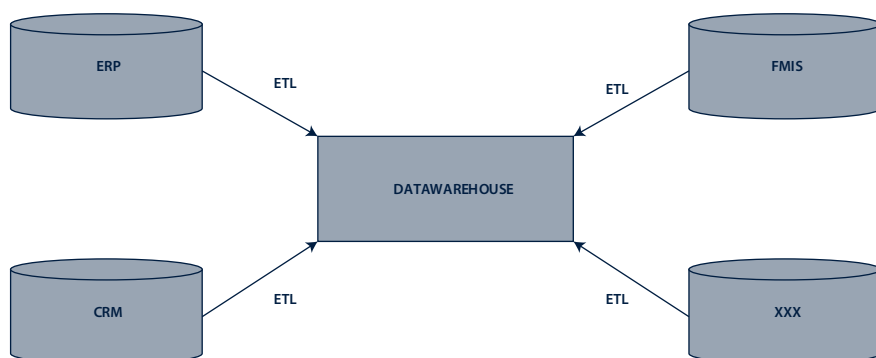
In een datawarehouse zitten gegevens uit allerlei verschillende databronnen. Welke data zitten in een datawarehouse, hoe wordt dat gebouwd en hoe kunnen we analyses maken op die data?

Opbouw datawarehouse

In ons vereenvoudigde datamodel is voor de overzichtelijkheid maar een aantal velden per tabel opgenomen. In professionele databases bevat een tabel tientallen velden. Ook het aantal tabellen is in het voorbeeld voor de overzichtelijkheid klein gehouden. Het aantal tabellen in professionele databases loopt snel op. Neem bijvoorbeeld het ERP-pakket Business Central van Microsoft voor het midden- en kleinbedrijf. Met die software kun je de volledige bedrijfsvoering verwerken: inkoop, verkoop, productie, voorraadbeheer, financiële admi-

nistratie, personeelszaken (human resource management – HRM), enzovoort. Het pakket heeft meer dan duizend tabellen en tabel *Klant* bevat al meer dan vijftig velden. Een vergelijkbaar softwarepakket voor het grootbedrijf, bijvoorbeeld SAP, heeft al meer dan tweeduizend tabellen. Het principe blijft overigens precies hetzelfde. Databases waarin de dagelijkse gang van zaken wordt opgeslagen en gemuteerd worden *Online Transaction Processing* (OLTP) systemen genoemd. Transacties kunnen zijn een inkooporder invoeren, een behandeling van een patiënt inboeken, een rechtszaak op de rol zetten of een order plaatsen.

Als je al die tabellen zou moeten raadplegen, kun je door de bomen het bos niet meer zien. Voor Data Analytics heb je nooit alle tabellen nodig, vaak maar een paar. Alleen die tabellen worden geladen in het datawarehouse en bovendien per tabel maar een beperkt aantal velden. Voor Data Analytics is bijvoorbeeld een adres, telefoonnummer en e-mailadres van een klant niet interessant. Stel dat je per klant wilt weten wat de omzet van dit jaar is vergeleken met de omzet van vorig jaar. Je toont klantnaam, omzet, omzet vorig jaar en procentueel verschil. Specifieke klantgegevens heb je daar niet voor nodig. De gegevens die je nodig hebt voor je data-analyses worden overgezet naar een datawarehouse. Binnen organisaties wordt met meerdere informatiesystemen gewerkt. Er wordt bijvoorbeeld gebruikgemaakt van een klantenvolgsysteem (Customer Relationship Management – CRM) en een systeem waarmee facilitaire processen gestroomlijnd kunnen worden (Facility Management Information System – FMIS). Daarnaast kun je nog gegevens nodig hebben van het internet, bijvoorbeeld demografische gegevens, gegevens van de Kamer van Koophandel of specifieke gegevens uit een Excelbestand. Alle informatie die je nodig hebt voor je data-analyse wordt dan overgezet naar het datawarehouse, zie figuur 1.8.



Figuur 1.8

In Power BI kunnen veel databronnen worden ingelezen, maar de data moeten altijd in tabellen staan. Dus ook als je data uit een webpagina haalt, zoekt Power BI alleen naar tabellen binnen deze webpagina. Dat geldt dus ook voor PDF-bestanden of Excelbestanden. Dit gaan we in volgende hoofdstukken oefenen.

Extract, Transform en Load proces

De verschillende gegevensbronnen selecteren, het transformeren van de gegevens en het inladen in het datawarehouse wordt aangeduid als Extract, Transform en Load, het ETL-proces, zie figuur 1.8. Bij transform kunnen de tabellen aangepast worden vóórdat ze geladen worden in het datawarehouse. Je wenst bijvoorbeeld alleen data vanaf het jaar 2024. In het bestand van de Kamer van Koophandel zijn bijvoorbeeld de eerste drie rijen niet gewenst omdat daarin adresgegevens van de KvK staan. Kolommen van de tabellen kun je verwijderen, toevoegen of aanpassen. Je wilt bijvoorbeeld je klanten segmenteren naar postcodegebieden. Je hebt daarvoor alleen de vier cijfers uit de postcode nodig en niet de gehele postcode, inclusief de twee letters. Je slaat in het datawarehouse dan niet bij iedere klant zijn postcode op, maar alleen de eerste vier cijfers van de postcode. Als alle gewenste gegevens zijn geselecteerd en eventueel aangepast (transform), kunnen de gegevens worden ingelezen (load) in het datawarehouse. Het ETL-proces gaan we ook in Power BI tegenkomen en oefenen.

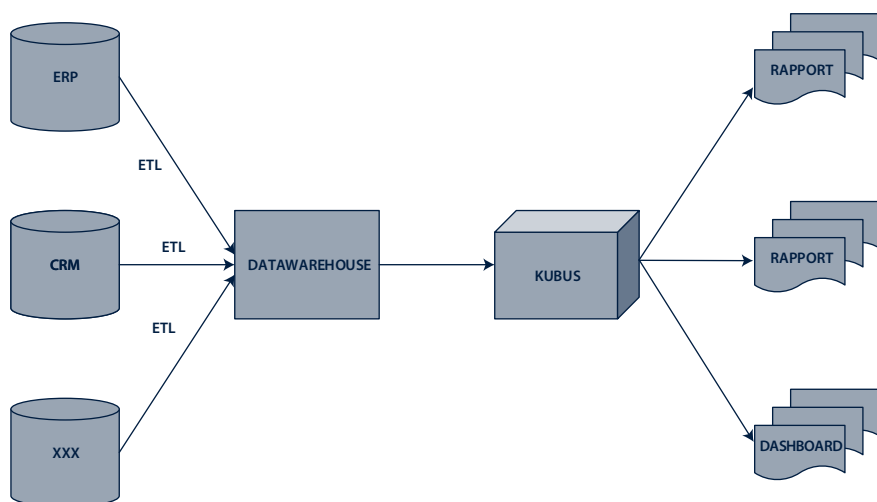
Analytische processen in een datawarehouse

In een datawarehouse worden de gegevens op een andere manier opgeslagen dan in relationele databases, die zijn beschreven in paragraaf 1.4 *Databases en tabellen* en 1.5 *Relaties tussen tabellen in een database*. Hoe dit gedaan wordt valt buiten het bestek van dit boek, maar de data worden geschikt gemaakt om er data-analyses op los te laten. De data worden opgeslagen in kolommen en die worden gecomprimeerd om de data zo klein mogelijk te maken. De data worden zodanig opgeslagen dat ze geschikt zijn voor analytische doeleinden. Bij data-analyses worden gegevens (bijna) altijd geaggregeerd (samengevat). Bijvoorbeeld de afzet per artikel, de aantallen verkochte eenheden worden dan samengevat (geaggregeerd) tot een totaal aantal. Het woord ‘per’ zit vaak in de gewenste informatie, bijvoorbeeld de omzet van het afgelopen jaar, per land, per regio. De omzet wordt dan per land en per regio geaggregeerd (samengevoegd) of geteld. De bekendste (aggregatie)functies die daarbij gebruikt worden zijn: *SUM()*, *AVERAGE()*, *MIN()* en *MAX()*, de bekendste telfuncties zijn *COUNT()* en *COUNTROWS()*. De processen die worden toegepast op het datawarehouse worden aangeduid als Online Analytical Processing (OLAP).

Merk nog op dat het niet altijd nodig is om zelf het datawarehouse te bouwen. Bij veel bedrijven bestaat zo’n datawarehouse al, bijvoorbeeld een Microsoft SQL Server Analysis Services-database (SSAS). Die kun je dan uiteraard direct gebruiken.

1.7 Kubussen

Kubussen is een bekend begrip binnen datawarehousing. Vanuit het datawarehouse worden kubussen met data gemaakt waarmee de gewenste rapporten en dashboards gemaakt worden, zie figuur 1.9.



Figuur 1.9

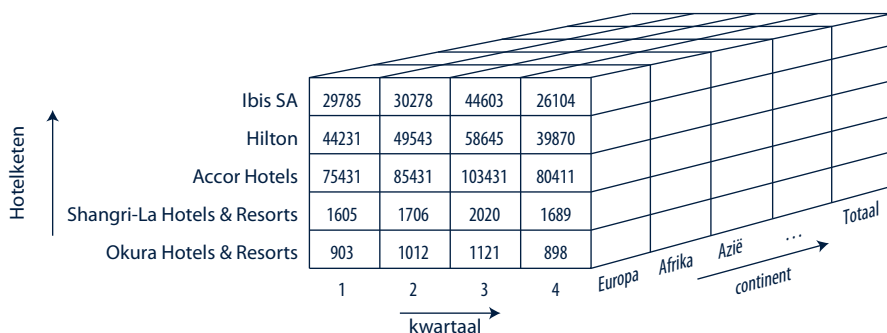
Een kubus kan gezien worden als een uitbreiding van een tweedimensionaal overzicht, in Excel een draaitabel genoemd. We nemen nog even het vereenvoudigde voorbeeld van Booking.com. Hotels zijn verbonden aan een moedermaatschappij, meestal een hotelketen, zoals Hilton, Sheraton, Holiday Inn of Ibis. Zo heeft de Hilton-groep wereldwijd meer dan 2200 hotels. Booking.com wenst per kwartaal, per hotelketen het aantal overnachtingen te weten. Die informatie kan tweedimensionaal getoond worden in een tabel, zie figuur 1.10.

Hotelketen	Kwartaal 1	Kwartaal 2	Kwartaal 3	Kwartaal 4
Ibis SA	178712	211946	267879	183421
Hilton	290732	360008	432898	310926
Accor Hotels	452135	609342	712443	476124
Shangri-La Hotels and Resorts	15412	17651	20186	15989
Okura Hotels & Resorts	8712	10412	11321	9011

Figuur 1.10

Dimensies en feiten

De assen van een tabel worden *dimensies* genoemd. Hierin hebben we de dimensie kwartaal en de dimensie hotelketen. In het snijpunt van de twee dimensies staan de *feiten*. Zo kan snel gezien worden dat in het tweede kwartaal bij de Hilton-groep het aantal overnachtingen 360.008 is (feit). De dimensies geven feiten een betekenis. Bijvoorbeeld, 360.008 aantal overnachtingen zegt op zich niets. Maar je geeft er betekenis aan door te zeggen dat het voor het tweede kwartaal voor de hotelketen Hilton is. Booking.com wenst dit verder uit te splitsen per continent. Er ontstaat nu een derde dimensie, continent. Met die derde dimensie ontstaat de kubus (eigenlijk is het een balk aangezien de dimensies niet even lang zijn), zie figuur 1.11.



Figuur 1.11

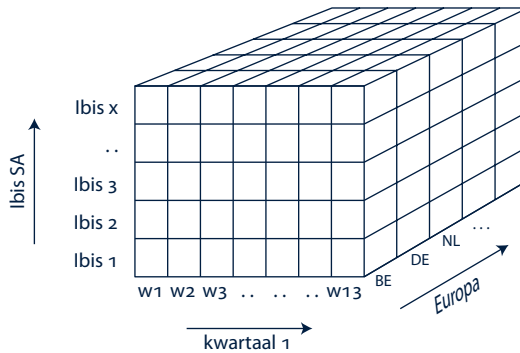
In al die blokjes in de kubus zitten de feiten (aantal overnachtingen). De kubus is op verschillende manieren te doorsnijden. Je kunt per continent doorsnijden maken, per kwartaal of per hotelketen. Als je in de dimensie continent een doorsnijding maakt op Europa, dan zie je het aantal hotelovernachtingen per kwartaal, per hotelketen van Europa. Merk op, dat in de laatste doorsnijding van de dimensie continent het totaal staat. Als je die doorsnijding toont, zie je weer figuur 1.10.

Slicing and Dicing

Als je in de dimensie hotelketen een doorsnijding maakt op Hilton, dan zie je het aantal overnachtingen per kwartaal per continent. Het doorsnijden van de kubus op verschillende dimensies wordt *Slicing and Dicing* genoemd. Dicing (kleine blokjes – kubussen), aangezien je ook een deelverzameling van de kubus kunt bekijken. Je hebt bijvoorbeeld de dealkubus (dice): alleen de eerste twee kwartalen van de hotelketens Hilton en Sheraton in de continenten Europa, Noord-Amerika en Zuid-Amerika.

Multi-dimensies

Je denkt nu wellicht, meer dimensies zijn er niet, want er bestaan maar drie dimensies (lengte, breedte, hoogte), alhoewel Albert Einstein al heeft aangetoond dat er een vierde dimensie, tijd, is. Maar binnen de wiskunde en het datawarehouse bestaan in theorie oneindig veel dimensies. Figuur 1.11 gaat over één jaar. Als je dezelfde gegevens over drie jaar wilt hebben, heb je drie van die kubussen naast elkaar, voor ieder jaar een kubus. Je hebt dan al vier dimensies: jaar, kwartaal, hotelketen en continent. Ook kun je nog extra dimensies toevoegen in de kubus zelf. Het aantal overnachtingen moet verder uitgesplitst worden per land, per hotel, per week, zie figuur 1.12. Hierin is kwartaal 1, Europa, Ibis SA uit figuur 1.11 verder opgesplitst in drie extra dimensies, weeknummer, land en hotelnaam.



Figuur 1.12

Zo kun je nog meer dimensies toevoegen, bijvoorbeeld per hotel, per kamer. Uit de kubussen kunnen de gewenste rapporten en dashboards gebouwd worden. Uiteraard zijn er ook meer feiten; zo kun je op dezelfde dimensies of andere dimensies de omzet (commissie) willen bepalen, of het aantal boekingen. Merk op dat ook één dimensie kan bestaan op de feiten, bijvoorbeeld alleen dimensie kwartaal; het totaal aantal overnachtingen in het vierde kwartaal.

1.8 Tabellen en relaties in een datawarehouse

Bekijk de video *Tabellen & relaties in Power BI*.



Tabellen

De dimensies en de feiten van een kubus worden weergegeven in tabellen. In figuur 1.13 staat het datamodel van ons vereenvoudigde voorbeeld van Booking.com.

Zoals te zien is, zijn de gegevens in de tabellen aangepast, bijvoorbeeld het adres van een klant zit niet meer in tabel *Klant*. Die informatie is voor managementinformatie niet nodig. Verder is een tabel *MetaSearch* toegevoegd. Voor boekingen wordt bijgehouden of direct via Booking.com is geboekt of via een zogenaamde meta-searcher, zoals Trivago of TripAdvisor. In tabel *MetaSearch* zitten alle meta-search bedrijven. De veldnamen worden in Power BI alfabetisch weergegeven in de tabellen.

We leven in een datatijdperk met een toenemende behoefte aan analyse van grote hoeveelheden data – data analytics – en visualisatie in grafische rapporten en dashboards. Dit wordt Business Intelligence genoemd. De BI-tools zijn in de loop der jaren sterk verbeterd. De eerste BI-tools konden alleen door specialisten gebruikt worden, maar Power BI is een tool voor de gebruikers: Self Service BI.

In dit boek worden eerst de onmisbare basisbegrippen uitgelegd, zoals gegevensbronnen, databases, relaties, datawarehouses, kubussen en big data, met veel voorbeelden en illustraties. Hierna start je met Power BI. De meest gebruikte visualisaties (visuals) en de DAX-taal worden toegelicht. Na het lezen van dit boek heb je een gedegen kennis van Power BI, die je direct in de praktijk kunt toepassen.

Het leren werken met een applicatie (app), zoals Power BI, kan alleen als je er hands-on mee aan de slag gaat. Om je hierbij te ondersteunen, maken we gebruik van blended learning. Ieder voorbeeld wordt voorafgegaan door een video-instructie op de online leeromgeving die bij het boek hoort. Na deze instructie ga je naar het boek. Daar wordt het voorbeeld stap-voor-stap toegelicht en voorzien van relevante schermafdrukken. Op de online leeromgeving zijn ook data en uitwerkingen te vinden. Daarnaast vind je er een introductievideo waarop je kunt zien dat miljoenen rijen aan data in fracties van seconden in Power BI tot een grafische weergave worden verwerkt.

In de derde druk zijn alle menu's herzien. Er zijn twee nieuwe visuals toegevoegd, de watervalgrafiek en de lintgrafiek. Een nieuwe paragraaf over de vensterfuncties WINDOW(), OFFSET() en INDEX() is opgenomen. Verder zijn sparklines toegevoegd, is het AI-hoofdstuk uitgebreid en zijn de data geactualiseerd.

Aan de slag met Power BI is een boek voor opleidingen binnen de ICT, management en logistiek. Ook is het geschikt voor zelfstudie.

Ben Groenendijk was werkzaam als docent informatica aan de Hogeschool Rotterdam. Hij is auteur van diverse boeken over databases en Excel.



9 789024 462698

www.boomstudent.nl
www.boomhogeronderwijs.nl