

# **AAN DE SLAG MET POWER BI**

Van data naar dashboard

Vierde editie

Ben Groenendijk

# **Boom**



Met onderstaande unieke activeringscode krijg je via **www.boomstudent.nl** toegang tot de online leeromgeving. Deze code is persoonsgebonden en gekoppeld aan de 4<sup>e</sup> editie. Na activering van de code is de online leeromgeving twee jaar toegankelijk. De code kan tot zes maanden na het verschijnen van een volgende editie geactiveerd worden. De code is eenmalig te gebruiken. Deze activeringscode is alleen bruikbaar voor een studentlicentie en geeft geen toegang tot de docentomgeving van het platform. Ben je docent? Dan heb je jouw activeringscode per e-mail ontvangen. Deze code activeer je op **boomdocent.nl**.

Omslagontwerp: Haagsblauw, Den Haag

© 2026 Groenendijk | 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. Auteursrecht ten aanzien van tekst- en datamining en machinelearning is nadrukkelijk voorbehouden. 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](http://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. No part of this publication may be reproduced in the context of text and data mining, machine learning or for any other purpose which is not expressly permitted by law without permission of Boom Uitgevers.*

ISBN 9789024474806

NUR 123

[www.boomstudent.nl](http://www.boomstudent.nl)

[www.boom.nl/hogeronderwijs](http://www.boom.nl/hogeronderwijs)

# Inhoud

|                                               |            |
|-----------------------------------------------|------------|
| <b>Voorwoord</b>                              | <b>5</b>   |
| <b>Inleiding</b>                              | <b>7</b>   |
| <b>1 Data modeling</b>                        | <b>11</b>  |
| 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         |
| <b>2 Power BI en visuals</b>                  | <b>33</b>  |
| 2.1 Power BI                                  | 33         |
| 2.2 Visuals                                   | 35         |
| 2.3 Visual berekeningen                       | 86         |
| 2.4 Publiceren                                | 101        |
| <b>3 Kunstmatige intelligentie</b>            | <b>105</b> |
| 3.1 Uitgevouwen structuur                     | 105        |
| 3.2 Belangrijkste beïnvloeders                | 112        |
| 3.3 Analyseren                                | 127        |
| <b>4 Extract, Transform en Load proces</b>    | <b>133</b> |
| <b>5 Kardinaliteit en filterrichting</b>      | <b>149</b> |
| 5.1 Een-op-een relaties                       | 149        |
| 5.2 Een-op-veel relaties                      | 154        |
| 5.3 Kruisfilterrichting bij relaties          | 159        |
| 5.4 Veel-op-veel relaties                     | 167        |

|          |                                                            |            |
|----------|------------------------------------------------------------|------------|
| <b>6</b> | <b>De meest gebruikte DAX-functies</b>                     | <b>175</b> |
| 6.1      | De dimensietabel Kalender                                  | 175        |
| 6.2      | Filteren met DAX                                           | 187        |
| 6.3      | Berekeningen in DAX met de functie <i>CALCULATE()</i>      | 190        |
| 6.4      | Kalendergebaseerde tijdintelligentie in DAX                | 204        |
| 6.5      | Klassenindeling in DAX                                     | 220        |
| 6.6      | Variabelen, cumulatieve waarden en Pareto                  | 225        |
| 6.7      | Vensterfuncties                                            | 233        |
| <br>     |                                                            |            |
| <b>7</b> | <b>Tips, trucs en nuttige visuals</b>                      | <b>257</b> |
| 7.1      | Metingen in één tabel plaatsen                             | 257        |
| 7.2      | Snelle metingen                                            | 260        |
| 7.3      | Tijdintelligentie met toekomstige datums                   | 263        |
| 7.4      | Metervisuals                                               | 266        |
| 7.5      | Infographic Designer                                       | 269        |
| 7.6      | DAX-queryweergave en TMDL-weergave                         | 273        |
| <br>     |                                                            |            |
| <b>8</b> | <b>Professionaliseren dashboard en rapport (Reporting)</b> | <b>277</b> |
| 8.1      | Bladwijzers                                                | 278        |
| 8.2      | Tooltips met detailinformatie                              | 282        |
| 8.3      | Drillthrough                                               | 287        |
| 8.4      | Knoppen en stuurtabelen                                    | 291        |
| 8.5      | Dynamische tekst en titels                                 | 300        |
| 8.6      | LeMair-dashboard                                           | 306        |
| <br>     |                                                            |            |
|          | <b>Over de auteur</b>                                      | <b>307</b> |
| <br>     |                                                            |            |
|          | <b>Index</b>                                               | <b>308</b> |

# 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 inleidende module 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 de eerste inleidende module 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](http://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](mailto:ben.groenendijk@outlook.com).

Februari 2026

Ben Groenendijk

# 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 module 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 module 2 ga je diverse visuals in Power BI maken. De benodigde bestanden en video-ondersteuning worden bij het boek geleverd via de online leeromgeving Boom Academie.

In module 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. Module 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 module 5 worden deze begrippen met concrete voorbeelden toegelicht.

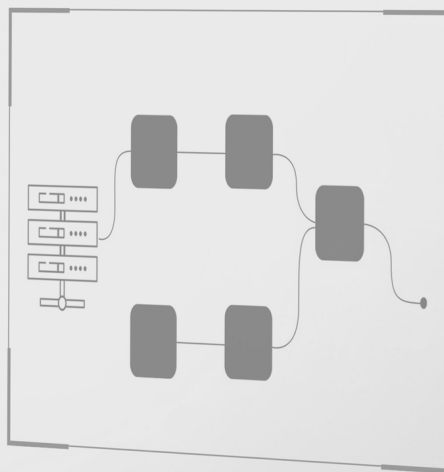
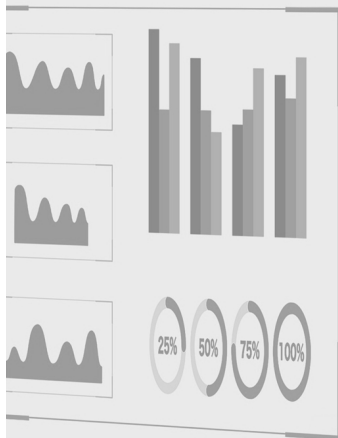
In module 6 wordt dieper ingegaan op de DAX-taal van Power BI en de datumtabel met tijdsintelligentie. De belangrijkste DAX-functies worden uitgebreid toegelicht. Een zeer belangrijke module.

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

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

### **Online leeromgeving Boom Academie**

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 deze module 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 modelleren van data noodzakelijk. Ook worden in deze module Big Data, Business Intelligence en Data Analytics toegelicht.

## **Bekijk de video:**

Introductie Power BI



## **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) managementinformatie 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, avonden 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, X, 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 in de Cloud of lokaal 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 een externe bron, 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 boekingssysteem 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 patiëntnummer, 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. | Aant. kamers | Commissie | Klantnr |
|-----------|---------|------------|-----------|--------------|-------------|--------------|-----------|---------|
| 931321100 | 678891  | 8-7-2025   | 10-7-2025 | Suite King   | 2           | 1            | \$ 45,30  | 100003  |
| 931321101 | 187612  | 31-12-2025 | 8-1-2025  | DZ standaard | 2           | 1            | \$ 213,80 | 564162  |
| 931321102 | 476324  | 17-9-2025  | 19-9-2025 | 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 zo veel 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 zo iets 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  | 8-7-2025 | 10-7-2025 | 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  | 3-2-2023   | 17-2-2023 | Royal Twin | 2           | 1            | \$ 264,90 | 100003  |
| 931321100 | 678891  | 8-7-2025   | 10-7-2025 | Suite King | 2           | 1            | \$ 45,30  | 100003  |
| 931324241 | 231765  | 28-12-2025 | 2-1-2026  | 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 3,4 miljoen hotels/appartementen/B&B's die geboekt kunnen worden, en het aantal boekingen wereldwijd en dus rijen in de gegevensgroep *Booking* 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 Dynamics 365 Business Central van Microsoft voor het midden- en kleinbedrijf. Met die software kun je de volledige bedrijfsvoering verwerken: inkoop, verkoop, productie, voorraadbeheer, financiële administratie, personeelszaken (human resource management – HRM), enzovoort.

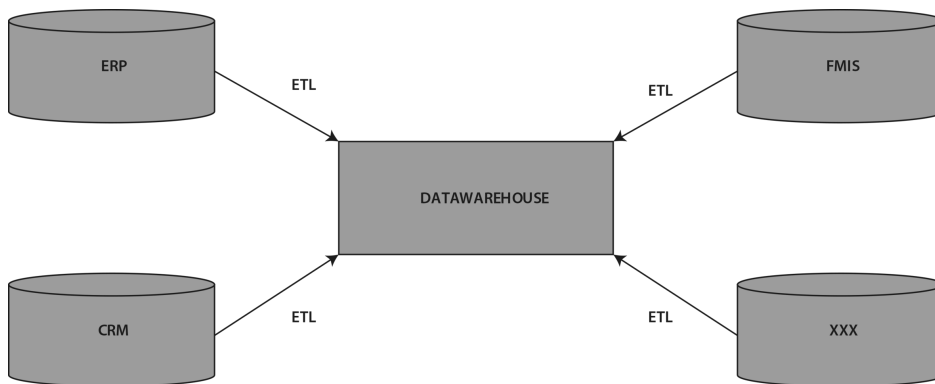
Het pakket heeft meer dan duizend tabellen en tabel *Klant* bevat al meer dan vijftig velden. Een vergelijkbaar softwarepakket voor het grootbedrijf, bijvoorbeeld SAP, heeft al meer dan tweeduizend tabellen. Het principe blijft overigens precies hetzelfde. Databases waarin de dagelijkse gang van zaken 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. Naast het ERP-pakket wordt er 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 modules 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 2026. 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.