

I DATABASE MANAGEMENT

1.1 Inleiding

Excel wordt veel gebruikt om met lijsten te werken. Meestal om informatie, zoals klantgegevens, op te slaan. Ook worden lijsten vaak gebruikt om gegevens te verwerken die uiteindelijk in een rapport worden opgenomen. In dat geval kun je beter spreken van het werken met een database dan het werken met een lijst. In dit boek spreken we in het vervolg dan ook over databases.

1.2 Excel als database

Excel kan prima functioneren als database, mits je de volgende beperkingen in acht neemt.

- Het aantal records of rijen kan nooit groter zijn dan 1.048.576.¹
- Het aantal velden van een record kan nooit meer zijn dan 16.384.

Zul je ooit met een groter aantal moeten werken, dan is Excel daar niet geschikt voor en zul je een ander pakket moeten kiezen dat dienst kan doen als database. Dat wil niet zeggen dat een database met meer dan 1.000.000 records een goed alternatief is om voor te kiezen. Geheugen en rekenkracht vormen, indien niet voldoende aanwezig, immers ook een bottleneck.

Als richtlijn, en meer ook niet, is een maximum van 250.000 records (rijen) en 32 velden (kolommen) aan te raden.

1.3 Aandachtspunten bij het opbouwen van een database

Een spreadsheet maken omvat meer dan alleen het invoeren van gegevens en het opstellen van formules. Een goed ontwerp vormt de basis voor succes. Zeker als

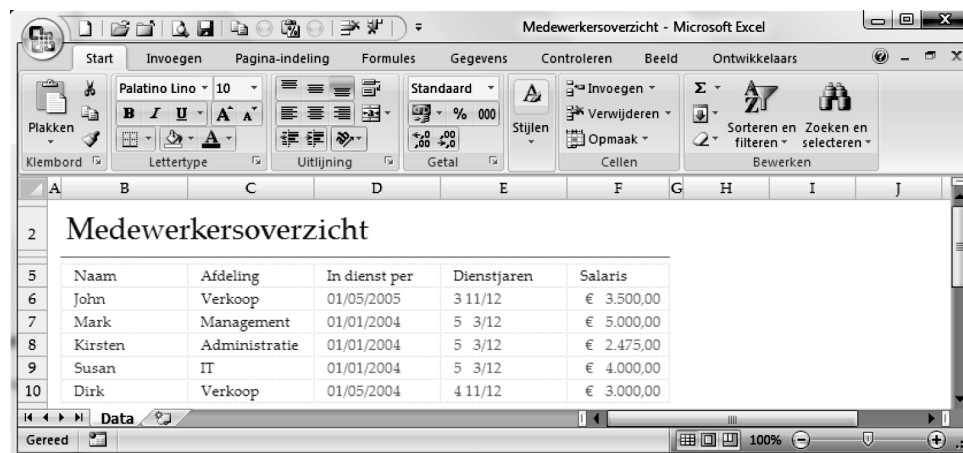
¹ In Excel versie 2003 bedraagt het maximum aantal records 65.536.

je werkt met grote databases, is het cruciaal om de volgende punten in acht te nemen.

- Gebruik heldere omschrijvingen voor elk titelveld of elke kolomkop in de eerste rij (hoofdrij) van de database.
- Gebruik slechts één rij als hoofdrij.
- Laat geen lege rij ontstaan; dat breekt de database in tweeën.
- Formateer elke kolom zodat ieder veld dezelfde opmaak heeft.
- Zorg dat elke kolom dezelfde soort informatie bevat; maak geen mix van datums en tekst.
- Bevries de eerste rij zodat de kolomkoppen altijd zichtbaar zijn.
- Sorteer een database altijd in een logische volgorde.

Door deze richtlijnen te volgen krijg je een overzichtelijke structuur en zijn eenvoudige maar ook complexere bewerkingen simpel uit te voeren.

Figuur 1.1 Kleine, goed gestructureerde database



The screenshot shows a Microsoft Excel window titled 'Medewerkersoverzicht - Microsoft Excel'. The ribbon includes 'Start', 'Invoegen', 'Pagina-indeling', 'Formules', 'Gegevens', 'Controleren', 'Beeld', and 'Ontwikkelaars'. The 'Gegevens' ribbon is active, showing options like 'Invoegen', 'Verwijderen', 'Opmaak', 'Sorteren en Zoeken en filteren', and 'Bewerken'. The worksheet contains a table with the following data:

	Naam	Afdeling	In dienst per	Dienstjaren	Salaris
6	John	Verkoop	01/05/2005	3 11/12	€ 3.500,00
7	Mark	Management	01/01/2004	5 3/12	€ 5.000,00
8	Kirsten	Administratie	01/01/2004	5 3/12	€ 2.475,00
9	Susan	IT	01/01/2004	5 3/12	€ 4.000,00
10	Dirk	Verkoop	01/05/2004	4 11/12	€ 3.000,00

1.4 Databases bewerken – Gegevens importeren

Gegevens vormen de basis voor alles wat je met Excel doet. De eerste stap in het maken van een spreadsheet, zeker als die als database moet dienen, heeft dan ook te maken met importeren van gegevens op een werkblad. Gegevens kunnen overal vandaan komen, maar tekstbestanden die uit datawarehouses worden geëxporteerd, vormen de grootste groep. Excel kent twee soorten tekstbestanden die je kunt importeren; de fixed length en delimited files.

Fixed length files

Fixed length files zijn tekstbestanden waarbij elk veld van een record een vaste lengte heeft. Als de tekst in het veld meer karakters bevat dan die vaste lengte, wordt de tekst afgekapt. Als de tekst in het veld minder karakters bevat dan die vaste lengte, wordt het veld aan het einde opgevuld met spaties.

Dit laatste geeft meteen aan waar je heel goed op moet letten als je met een fixed length tekstbestand te maken krijgt. Om de juiste waarden te kunnen vinden, of om een goede analyse te kunnen maken, is het bij een fixed length tekstbestand essentieel dat je alle overbodige spaties in alle velden verwijdert.

Een voorbeeld maakt dit duidelijk. Stel dat je in een lijst met automerken een uitgebreid filter gebruikt waarbij je als criterium de merknaam BMW gebruikt. De meeste automerken hebben een naam die langer is dan drie karakters, maar nooit meer dan 15. De merknaam BMW wordt in alle velden dus opgenomen met 12 extra spaties. In je zoektocht naar welke statistische waarde dan ook, zul je in deze opzet geen enkele treffer vinden, omdat de merknaam BMW gewoon niet voorkomt.

De merknaam BMW komt echter wel voor, maar je kunt natuurlijk niet verwachten dat er iemand is die deze naam als criterium zal gebruiken.

De consequentie is helder: op de een of andere wijze zul je alle overbodige spaties uit alle velden moeten wissen. De vraag die rest is: hoe?

De functie Spaties.Wissen

Gelukkig beschikt Excel over de functie Spaties.Wissen. Met deze functie kun je alle voorloop- en vervolgs spaties uit de tekst verwijderen, terwijl spaties tussen de woorden blijven staan. Je kunt deze functie gebruiken in een extra kolom en het resultaat van ieder veld uit die kolom kopiëren naar het veld in de originele kolom. De functie Spaties.Wissen vind je in de functiebibliotheek op het tabblad Formules onder de groep Tekst.

Figuur 1.2 De functie *Spaties.Wissen*



Uiteraard zal het voor grote databases een hele opgave zijn om alle velden op deze wijze op te schonen. Het is echter beter dan werken met onjuiste gegevens.

Delimited files

Delimited files zijn tekstbestanden waarvan de velden worden gescheiden door speciale karakters. De meest gebruikte karakters zijn de tab en de komma, maar je kunt ieder willekeurig karakter gebruiken. Delimited files zijn moderner dan fixed length files en meestal afkomstig van een pc of server. Ze zijn vaak kleiner in omvang dan de fixed length files en ook eenvoudiger te bewerken.

Essentieel bij het werken met delimited files is dat de persoon die het bestand opslaat, als scheidingsteken een karakter kiest dat niet voorkomt in de velden met gegevens. Indien dit namelijk wel zo, is komt het scheidingsteken te vaak binnen één record voor en weet het importprogramma niet meer waar het zich in het proces bevindt.

Het zal duidelijk zijn dat het gebruiken van een komma als scheidingsteken niet de beste keuze is. Het Pipe-teken (|) is dat in mijn beleving wel, omdat dit teken nooit in een woord of een getal zal voorkomen.

Tekstbestanden importeren

Zowel fixed length files als delimited files kun je eenvoudig importeren. In beide gevallen werkt dat op dezelfde wijze.

- Klik op de Office-knop en kies Openen
- Geef in het filter voor bestandstypen aan dat je een tekstbestand wilt selecteren

- Selecteer het bestand dat je wilt importeren
- Klik op Openen, zodat de dialoog Wizard Tekst importeren geopend wordt

Figuur 1.3 Wizard Tekst importeren – Het bestandstype kiezen



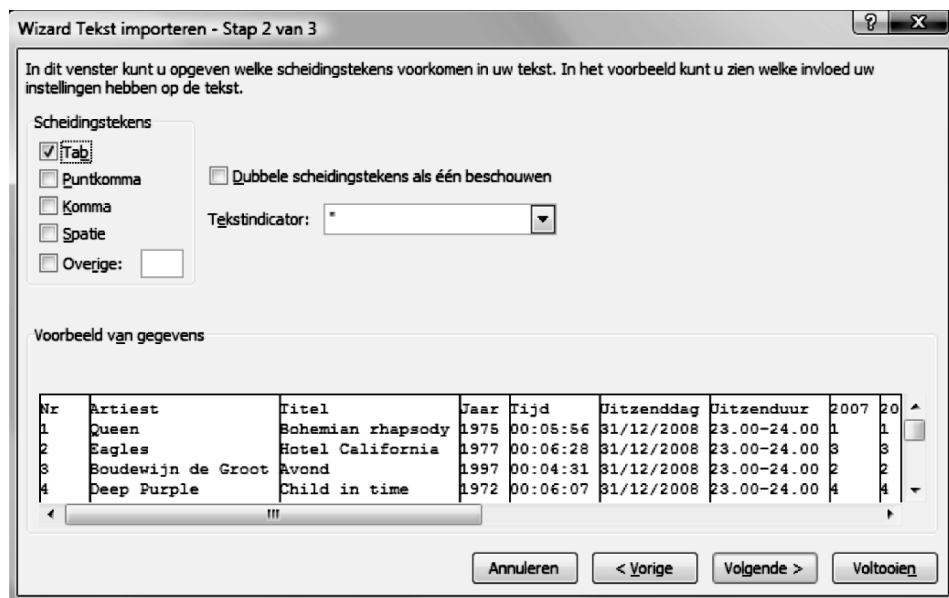
De eerste stap in het proces is het kiezen van het bestandstype.

In dit voorbeeld wordt een tekstbestand van het type Gescheiden (delimited) geïmporteerd, waarbij de velden zijn gescheiden door het Tab-teken. Misschien niet de beste keuze, maar helaas wel de keuze die het meeste voorkomt.

In de eerste stap van dit proces kun je aangeven bij welke regel het importproces moet beginnen. In dit voorbeeld is dat rij 1 (de rij met veldnamen) maar die rij kun je overslaan door bij Importeren starten bij rij een 2 in te vullen.

Klik op Volgende > voor de tweede stap in het proces. Hier kun je aangeven welk scheidingsteken er is gebruikt. In dit voorbeeld is dat een Tab. Zou je het Pipe-teken (|) gebruiken, dan vul je dat teken bij Overige in en haal je het vinkje bij Tab weg.

Figuur 1.4 Wizard Tekst importeren – Een scheidingsteken kiezen



Klik op Volgende > voor de derde stap in het proces.

Figuur 1.5 Wizard Tekst importeren – De opties voor velden aangeven

Wizard Tekst importeren - Stap 3 van 3

In dit venster kunt u voor elke kolom het gegevenstype instellen.

Gegevenstype per kolom

☒ Standaard
☐ Tekst
☐ Datum:
☐ Kolom overslaan bij importeren

De optie Standaard converteert numerieke waarden naar getallen, datumwaarden naar datums en alle overige waarden naar tekst.

Voorbeeld van gegevens

Stand	Standaard	Standaard	Stand	Standaard	Standaard	Standaard	Stand	St
02	Artiest	Titel	Jaar	Tijd	Uitzenddag	Uitzenduur	2007	20
1	Queen	Bohemian rhapsody	1975	00:05:56	31/12/2008	23.00-24.00	1	1
2	Eagles	Hotel California	1977	00:06:28	31/12/2008	23.00-24.00	3	3
3	Boudewijn de Groot	Avond	1997	00:04:31	31/12/2008	23.00-24.00	2	2
4	Deep Purple	Child in time	1972	00:06:07	31/12/2008	23.00-24.00	4	4

Stap drie is de interessantste in het proces. Voor ieder veld kun je aangeven of je het veld wilt importeren en welke opmaak dit veld moet hebben. Dit laatste is natuurlijk handig bij zeer grote bestanden waarvan je weet dat je maar een heel klein deel van de kolommen zult gebruiken.

Klik op Voltooien als je alle opties hebt doorlopen en je zeker weet dat je alle velden hebt geselecteerd die je wilt importeren. Het tekstbestand wordt dan in Excel opgenomen.

Figuur 1.6 Een door komma's gescheiden tekstbestand in Excel geïmporteerd

The screenshot shows the Microsoft Excel interface with a CSV file named 'Top 2000 van 2008 Tekst' imported. The data is organized into a table with the following columns: Nr, Artiest, Titel, Jaar, Tijd, Uitzendde, Uitzenddu, 2007, 2006, 2005, 2004, 2003. The rows contain data for various artists and their chart positions over time.

Nr	Artiest	Titel	Jaar	Tijd	Uitzendde	Uitzenddu	2007	2006	2005	2004	2003
1	Queen	Bohemian	1975	00:05:56	#####	23.00-24.0	1	1	2	1	1
3	Eagles	Hotel Cali	1977	00:06:28	#####	23.00-24.0	3	3	3	2	2
4	Boudewijn	Avond	1997	00:04:31	#####	23.00-24.0	2	2	1	5	8
5	Deep Pur	Child in ti	1972	00:06:07	#####	23.00-24.0	4	4	4	3	3
6	Led Zeppe	Stairway t	1971	00:08:00	#####	23.00-24.0	5	5	5	4	4
7	Meat Loaf	Paradise t	1978	00:08:24	#####	23.00-24.0	10	7	6	6	5
8	Animals	The house	1964	00:04:26	#####	23.00-24.0	7	12	8	8	7
9	Alan Pars	Old and w	1982	00:04:53	#####	23.00-24.0	9	11	9	11	9
10	Guns n' R	Novembe	1992	00:08:50	#####	22.00-23.0	20	13	10	7	6
11	U2	One	1992	00:04:32	#####	22.00-23.0	8	6	17	14	82

Het enige wat je nog moet doen om met dit bestand een beetje prettig te kunnen werken is het instellen van de juiste kolombreedte voor elk veld of elke kolom. Het bestand is hiermee klaar voor gebruik en kan prima als database dienen.

1.5 Databases bewerken – Data entry

Als je werkt met databases, zul je vroeg of laat ook gegevens moeten bewerken. Meestal is dat geen probleem, maar ook voor de gevorderde gebruiker zijn er wat handigheidjes die de moeite waard zijn om op een rijtje te zetten.

- Om dezelfde waarden in een gegevensreeks in te voeren, selecteer je eerst de gegevensreeks, waarna je de invoer bevestigt door gelijktijdig op Control en Enter te drukken.
- Om een gegevensreeks te vullen met oplopende waarden, kun je de Control-toets ingedrukt houden en de vulgreep naar het onderste veld van de gegevensreeks slepen.
- Je kunt de vulgreep ook gebruiken om een serie oplopende datums in te vullen. Excel herkent namelijk ook maanden en dagen van de week.
- De vulgreep kun je ook met de rechtermuisknop slepen zodat een shortcut menu met handige extra's zichtbaar wordt.

Tip

Probeer bij het invoeren van gegevens ook eens de optie **Uit vervolgkeuzelijst selecteren...** Deze optie verschijnt in een shortcut menu door in een willekeurig record van de data-

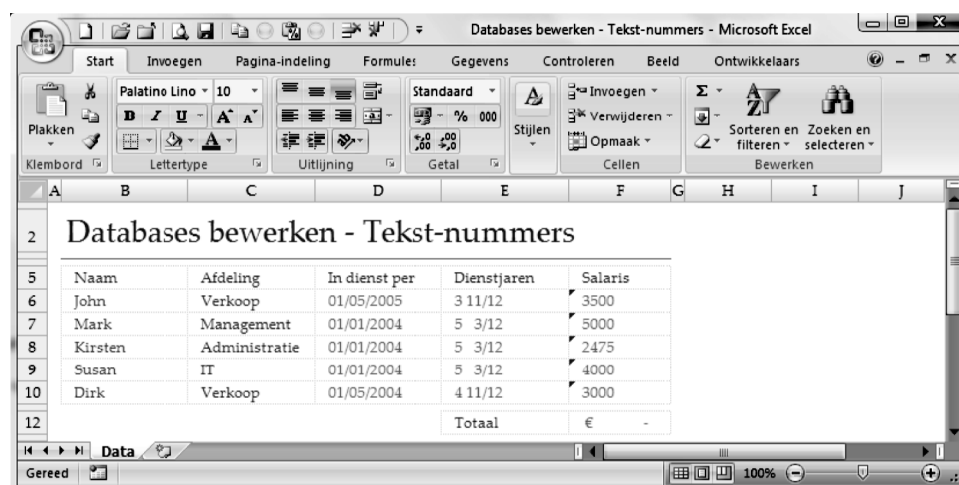
base, of de eerste lege regel onder de database, met de rechtermuisknop een veld aan te klikken.

1.6 Databases bewerken – Tekstnummers

Iedereen heeft wel eens dat groene driehoekje aan de linkerbovenkant van een cel gezien. Als je de betreffende cel selecteert en de muis over het uitroepteken beweegt, verschijnt de mededeling ‘Getal opgeslagen als tekst’.

Voor na het importeren van gegevens worden nummers vaak weergegeven als tekst. Veel mensen vinden dat groene driehoekje vervelend en menigeen negeert het. Die laatste groep is echter slecht af, want als je er niets mee doet, zullen veel bewerkingen met deze tekstnummers totaal de mist in gaan. Kijk maar naar de optelling in cel F12 in figuur 1.7.

Figuur 1.7 Getal opgeslagen als tekst



In cel F12 is de formule `=SOM(F6:F10)` ingevoerd. De uitkomst van deze formule zou € 17.975 moeten opleveren, maar er wordt echter € - weergegeven. Niet helemaal juist, maar hoe los je dit op?

Vaak wordt gedacht dat de klus wordt geklaard door het bereik nog een keer op te maken. Niets is echter minder waar.

- Selecteer het bereik F6:F10
- Druk op Ctrl + I, Tabblad Getal en Financieel als opmaak

Er verandert niets; alle cellen blijven opgemaakt als tekst. Je zult dus een extra handeling moeten verrichten.

Excel heeft voor de oplossing van dit probleem een aantal mogelijkheden. De simpelste methode is de cel selecteren en F2, gevolgd door Enter in te drukken. Dat werkt redelijk snel bij een klein aantal cellen, maar is niet effectief bij grotere bereiken.

Voor het bewerken van een groot bereik kun je het volgende doen:

- Kopieer een lege cel of een cel met de waarde o
- Selecteer het bereik dat moet worden bewerkt
- Kies op tabblad Start de optie Plakken > Plakken speciaal...
- Kies Plakken: = Waarden, Bewerking: = Optellen
- Klik OK

Figuur 1.8 De dialoog Plakken speciaal



De tekstnummers zijn nu allemaal omgezet naar getallen en de uitkomst in F12 is € 17.975. Precies dat wat je zou verwachten.

De reden waarom deze methode werkt, is dat een rekenkundige bewerking op een reeks tekstnummers als resultaat een reeks getallen oplevert.

Let op!

Getallen die zijn opgemaakt als tekst zijn te herkennen aan het feit dat ze standaard links zijn uitgelijnd.

1.7 Databases bewerken – Lege velden vullen

Na het importeren van gegevens uit een externe bron kan het voorkomen dat je met een overzicht wordt opgezadeld dat er uitziet zoals in figuur 1.9. Het op deze wijze opmaken van een rapport is niet ongebruikelijk en zal voor weinig verwarring zorgen. Het zal voor iedereen namelijk duidelijk zijn dat een naam in kolom F betrekking heeft op meerdere records.

Figuur 1.9 Een database met lege velden



	Datum	Bedrag	Aantal	Betaalmiddel	Behandeld door	Regio	Type klant
6	02/01/2009	€ 3.000,00	117	Automatisch	Claudia	Midden	Bestaand
7	02/01/2009	€ 4.000,00	131			Overig Nederland	
8	02/01/2009	€ 13.000,00	145	Contant		Midden	
9	02/01/2009	€ 50.000,00	119			Overig Nederland	Nieuw
10	02/01/2009	€ 15.000,00	395			Westland	
11	02/01/2009	€ 6.000,00	215	Op rekening		Midden	Bestaand
12	02/01/2009	€ 2.878,00	286			Overig Nederland	
13	02/01/2009	€ 65.000,00	248			Westland	
14	02/01/2009	€ 240,00	143	Automatisch	Yvonne	Midden	
15	02/01/2009	€ 6.762,00	146	Op rekening		Midden	

Het bewerken van een dergelijk rapport levert echter wel problemen op. Als je dit rapport moet sorteren, zorgen de lege cellen voor een hoop ellende en ben je daarna niet meer in staat om te vertellen wie nu precies wat verkocht.

Indien dit probleem een kleine lijst betreft, is het eenvoudig om de ontbrekende waarden te kopiëren of handmatig in te voeren.

Je kunt ook dubbelklikken op de vulgreep van een gevuld veld, maar dat werkt alleen als de onderliggende velden links of rechts aansluiten op een ander veld. In figuur 1.9 werkt dit dus niet bij cel F6, maar wel bij cel H6, H9 en H11.

Als je met grotere lijsten of databases werkt, ben je al dat geklik natuurlijk al snel zat en heb je een iets efficiëntere manier nodig om alle lege velden op te vullen. De volgende methode biedt uitkomst.

- Selecteer het bereik dat de lege velden bevat (E6:E15)
- Kies op tabblad Start de optie Bewerken > Zoeken & Selecteren > Selecteren speciaal...
- Klik in de dialoog Selecteren speciaal op de optie Lege waarden
- Klik OK

Figuur 1.10 De dialoog Selecteren speciaal



Van het bereik dat je in eerste instantie hebt geselecteerd zijn nu alleen de lege cellen nog actief.

- Voer de volgende formule in de formulebalk in: =E6
- Druk tegelijk op Ctrl + Enter

Alle lege velden zijn nu gevuld met een formule.

- Selecteer het bereik opnieuw