

Voorbeeld normaliseren

Voor het bepalen van de indeling van de database bestaan verschillende technieken. In deze tutorial zal ik de techniek **Normaliseren**, die in de jaren 70 door de Amerikaan Edgar Codd is opgesteld, uitleggen.

Als je een database ontwikkelt (en dit gaat dan niet alleen over je MySQL-database voor je PHP-website) is het belangrijk dat het aan een aantal eisen voldoet. Eén daarvan is de eis dat het bestand optimaal ontworpen is. Dit betekent dat er geen overbodige *redundantie* mag zijn.

Er zijn meerdere stappen om naar deze goede database te komen. Deze stappen zijn Normaalvormen. Dit zijn de Normaalvormen waarmee ik in deze tutorial ga werken:

0NV (Nulde Normaalvorm - Inventarisatie)

1NV (Eerste Normaalvorm)

2NV (Tweede Normaalvorm)

3NV (Derde Normaalvorm)

Verder zijn er ook nog de BCNV (Boyce Codd Normaalvorm), 4NV en 5NV. Op deze laatste Normaalvormen ga ik niet in omdat deze bijna nooit gebruikt worden.

Uitgangspunt voor het normaliseren is steeds de informatiebehoefte van de toekomstige gebruiker van de database. De indeling en inhoud van de tabellen wordt bepaald door de informatie die de gebruiker wenst te zien.

Nulde Normaalvorm

Om tot de Nulde Normaalvorm te komen moeten we de informatiebehoefte gaan inventariseren. Voor het gemak heb ik even een afleverbon gemaakt:

AFLEVERBON				
ORDERNR.: 3405		ORDERDATUM: 15-06-2004		
KLANTNR.: 100658				
C1000 Oud-Beijerland Graaf van Egmondstraat 67B 3261 AK Oud-Beijerland				
Artnr.	Art.omschrijving	Aantal	Prijs	Totaal
15201	Biologisch ei 6st per 20	1	12,50	12,50
21987	Biologische kaas Campina 10*0.5 kg Aanbiedingpr.	6	8,99	53,94
12365	Rolcontainer	1	125,00	125,00
12366	CBL Krat Zwart 24, emballage	7	3,60	25,20
				Totaal € 216,64

Nu gaan we inventariseren. Dit betekent dat we alle gegevens op de afleverbon netjes onder elkaar gaan zetten. We krijgen dan het volgende:

ONV
ORDERS
ordernr
orderdatum
klantnr
klantnaam
adres
postcode
plaats
artnr
artomschrijving
aantal
prijs
regeltotaal
eindtotaal

Zoals je ziet heb ik de inventarisatie ook een naam gegeven. Deze staat in hoofdletters.

Eerste Normaalvorm

Na het inventariseren is het de bedoeling dat we naar de Eerste Normaalvorm moeten gaan. Dit doen we altijd met de volgende stappen:

1. Verwijder alle procesgegevens.
2. Geef de sleutel van de groep aan.
3. Geef de deelverzameling aan die een herhaald aantal keren voorkomt t.o.v. de primaire sleutel.
4. Herhaal de sleutelgegevens van de oorspronkelijke groep samen met de gegevens van de zich herhalende deelverzameling als een nieuwe groep.
5. Verwijder de zich herhalende deelverzameling uit de oorspronkelijke groep.

Oké, dan nu een voorbeeld:

AFLEVERBON				
ORDERNR.: 3405		ORDERDATUM: 15-06-2004		
KLANTNR.: 100658				
C1000 Oud-Beijerland Graaf van Egmondstraat 67B 3261 AK Oud-Beijerland				
Artnr.	Art.omschrijving	Aantal	Prijs	Totaal
15201	Biologisch ei 6st per 20	1	12,50	12,50
21987	Biologische kaas Campina 10*0.5 kg Aanbiedingpr.	6	8,99	53,94
12365	Rolcontainer	1	125,00	125,00
12366	CBL Krat Zwart 24, emballage	7	3,60	25,20
				Totaal € 216,64

1. Verwijder alle procesgegevens.

In mijn voorbeeld is het duidelijk wat de procesgegevens zijn: Het regeltotaal wordt berekend uit aantal

* prijs en het eindtotaal wordt berekend als som van alle regeltotalen. Deze twee strepen we dus weg.

2. Geef de sleutel van de groep aan.

Een sleutel is altijd uniek. Je ziet ze op het Internet vaak als 'id'. Het kan natuurlijk anders heten, maar het heeft wel dezelfde functie: het uniek identificeren van een record (=rij).

In ons voorbeeld is de sleutel ordernr, immers een klant kan meerdere keren een order plaatsen, maar de orders kunnen nooit hetzelfde nummer hebben.

De sleutel geven we aan door het te onderstrepen.

3. Geef de deelverzameling aan die een herhaald aantal keren voorkomt t.o.v. de primaire sleutel.

Deze deelverzameling noemt men ook wel eens de Repeterende Groep (RG). Dit zijn de gegevens die vaker voorkomen. In ons voorbeeld is dit het tabelletje met de bestelde goederen. Geef deze gegevens aan in je inventarisatie. We hebben nu dus het volgende:

Normaliseren

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

RG artnr

RG artomschrijving

RG aantal

RG prijs

X regeltotaal (procesgegeven)

X eindtotaal (procesgegeven)

4. Herhaal de sleutelgegevens van de oorspronkelijke groep samen met de gegevens van de zich herhalende deelverzameling als een nieuwe groep.

Kijk maar naar het voorbeeld:

ONV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

RG artnr

RG artomschrijving

RG aantal

RG prijs

BESTELDE_ARTIKELEN

ordernr

artnr

artomschrijving

aantal

prijs

Zoals je ziet heeft de tweede groep nog geen sleutel. De sleutel moet je zodanig kiezen dat er zo min mogelijk herhaalde groepen voorkomen t.o.v. deze sleutel. Het liefst natuurlijk geen herhaalde groepen

Normaliseren

meer, anders moet je de vorige stappen nog een keer herhalen.

Meestal kun je een combinatie nemen van de sleutel van de oorspronkelijke groep en het gegeven dat in de Repeterende Groep de sleutelrol vervult. De sleutel wordt in dit geval een combinatie van ordernr en artnr.

5. Verwijder de zich herhalende deelverzameling uit de oorspronkelijke groep.

Hier moeten we de RG dus weer opruimen en dan hebben we de Eerste Normaalvorm:

1NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

artnr

artomschrijving

aantal

prijs

Tweede Normaalvorm

Nu is het tijd om naar de Tweede Normaalvorm te gaan. Dit doen we ook weer met een aantal vaste stappen:

1. Geef de attributen aan die niet functioneel afhankelijk zijn van de volledige sleutel.
2. Formeer een aparte groep voor ieder deel van de sleutel waarvan de attributen functioneel afhankelijk zijn.
3. Neem in iedere groep de attributen met het bijbehorende sleuteldeel op en wijs de primaire sleutel aan.
4. Verwijder deze attributen uit de oorspronkelijke groep.

Het voorbeeld:

AFLEVERBON				
ORDERNR.: 3405		ORDERDATUM: 15-06-2004		
KLANTNR.: 100658				
C1000 Oud-Beijerland Graaf van Egmondstraat 67B 3261 AK Oud-Beijerland				
Art.nr.	Art.omschrijving	Aantal	Prijs	Totaal
15201	Biologisch ei 6st per 20	1	12,50	12,50
21987	Biologische kaas Campina 10*0.5 kg Aanbiedingpr.	6	8,99	53,94
12365	Rolcontainer	1	125,00	125,00
12366	CBL Krat Zwart 24, emballage	7	3,60	25,20
				Totaal € 216,64

1. Geef de attributen aan die niet functioneel afhankelijk zijn van de volledige sleutel.

We herhalen nog even de Eerste Normaalvorm:

1NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

Normaliseren

BESTELDE_ARTIKELEN

ordernr

artnr

artomschrijving

aantal

prijs

De eerste groep komt niet in aanmerking voor deze stap omdat het niet beschikt over een samengestelde sleutel. Binnen de andere groep zijn er wel gegevens die functioneel afhankelijk zijn van een deel van de sleutel.

Kijk maar eens naar artomschrijving en prijs. Blijkbaar zijn deze afhankelijk van artnr en niet van ordernr. Je kunt dit nagaan door te kijken wat er verandert als het artnr gewijzigd wordt.

We geven dit zo aan in het voorbeeld:

1NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

> artnr

> artomschrijving

aantal

> prijs

Let erop dat aantal niet alleen afhankelijk is van artnr. Het aantal per artikel kan verschillen per order.

2. Formeer een aparte groep voor ieder deel van de sleutel waarvan de attributen functioneel afhankelijk zijn.

Het kan gebeuren dat een samengestelde sleutel in meerdere delen gesplitst kan worden en dat van ieder deel afzonderlijk attributen functioneel afhankelijk zijn. Er moeten dan meerdere groepen gevormd worden. In het voorbeeld ontstaat slechts één nieuwe groep: ARTIKELEN.

3. Neem in iedere groep de attributen met het bijbehorende sleuteldeel op en wijs de primaire sleutel aan.

Normaliseren

1NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

> artnr

> artomschrijving

aantal

> prijs

ARTIKELEN

artnr

artomschrijving

prijs

4. Verwijder deze attributen uit de oorspronkelijke groep.

Na deze stap hebben we onze Tweede Normaalvorm en begint het al lekker op te schieten:

2NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

artnr

aantal

ARTIKELEN

artnr

artomschrijving

prijs

Derde Normaalvorm

De laatste Normaalvorm die ik behandel in deze tutorial is de Derde Normaalvorm. Hiervoor hebben we de volgende stappen:

1. Geef de niet-sleutel attributen aan die functioneel afhankelijk zijn van andere niet-sleutel attributen.
2. Formeer een aparte groep voor ieder attribuut of combinatie van attributen, waar andere attributen functioneel van afhankelijk zijn.
3. Neem in iedere groep de attributen met bijbehorende sleutel op en wijs de primaire sleutel aan.
4. Verwijder de attributen van de nieuwe groep(en) uit de oorspronkelijke groep.

Ik begin nog maar eens met de uitkomst van de Tweede Normaalvorm:

2NV

ORDERS

ordernr

orderdatum

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

artnr

aantal

ARTIKELEN

artnr

artomschrijving

prijs

1. Geef de niet-sleutel attributen aan die functioneel afhankelijk zijn van andere niet-sleutel attributen.
Als we kijken in de tabel ORDERS zien we dat er een aantal gegevens zijn die niet afhankelijk zijn van ordernr. Deze zijn klantnaam, adres, postcode en plaats. Deze zijn afhankelijk van het klantnr. Klantnr is in deze tabel een niet-sleutelattribuut, dus deze stap is makkelijk uit te voeren:
(de afhankelijkheid is aangegeven met een A)

Normaliseren

2NV

ORDERS

ordernr

orderdatum

A klantnr

A klantnaam

A adres

A postcode

A plaats

BESTELDE_ARTIKELEN

ordernr

artnr

aantal

ARTIKELEN

artnr

artomschrijving

prijs

2. Formeer een aparte groep voor ieder attribuut of combinatie van attributen, waar andere attributen functioneel van afhankelijk zijn.

Vrij vertaald moeten we voor de groep die we net aangegeven hebben een nieuwe groep maken:

2NV

ORDERS

ordernr

orderdatum

A klantnr

A klantnaam

A adres

A postcode

A plaats

KLANTEN

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

artnr

aantal

Normaliseren

ARTIKELEN

artnr

artomschrijving

prijs

3. Neem in iedere groep de attributen met bijbehorende sleutel op en wijs de primaire sleutel aan.

We moeten dus de sleutel waarvan de nieuwe groep afhankelijk is in de nieuwe groep plaatsen, daarna moeten we de sleutel aangeven. De nieuwe groep ziet er dan zo uit:

KLANTEN

klantnr

klantnaam

adres

postcode

plaats

Let er wel op dat klantnr niet zomaar uit ORDERS gehaald kan worden, dan zouden de orders namelijk nooit aan een klant gekoppeld kunnen worden.

4. Verwijder de attributen van de nieuwe groep(en) uit de oorspronkelijke groep.

Als we deze stap hebben gedaan zijn we klaar met het normaliseren. Dit ziet er dan zo uit:

3NV

ORDERS

ordernr

klantnr

orderdatum

KLANTEN

klantnr

klantnaam

adres

postcode

plaats

BESTELDE_ARTIKELEN

ordernr

artnr

aantal

Normaliseren

ARTIKELEN

artnr

artomschrijving

prijs

Afsluiting

In deze tutorial heb ik geprobeerd om je een klein beetje de beginselen van het Normaliseren onder de knie te brengen. Natuurlijk heb ik hier een makkelijk voorbeeld gebruikt. Maar met te complexe voorbeelden kom je er met een begin niet uit.

Ik hoop dat mensen gaan inzien dat normaliseren een geweldige manier is om tot een goede structuur van de database te komen.

Begrippen & Literatuur

Begrippen

- Redundantie: Het dubbel opslaan van gegevens.
- Gegevens: Feiten of gebeurtenissen die op een bepaalde manier vastgelegd zijn.
- Informatie: De betekenis die aan gegevens ontleend wordt.
- Procesgegevens: Gegeven wat uit een berekening van andere gegevens wordt afgeleid.
- Tupel: Rij in een tabel
- Attribuut: Kolom in een tabel
- Graad: Het aantal attributen in een tabel
- Kardinaliteit: Het aantal tupels in een tabel
- Functionele afhankelijkheid: We zeggen dat een gegeven B functioneel afhankelijk is van een ander gegeven A als er bij een waarde A één (en niet meer dan één) waarde van B kan voorkomen (naam van een artikel is functioneel afhankelijk van het artnr).

Literatuur

Oriëntatie op de informatieanalyse (I. Korpershoek en B. Groenendijk, Academic Service, ISBN 90 395 1628 6)

Bron: <http://www.phphulp.nl/php/tutorials/3/150/265/>