

Een breed toepasbare oplossing voor dimensionele modellen

Modelleren van plancijfers met variabel detailniveau

Mark Zwijsen

In een datawarehouse-omgeving kan men plan- en realisatiecijfers integreren, waarbij het detailniveau van de plancijfers variabel is. Op basis van generieke modelleringsconcepten kan een oplossing beschreven worden, die past binnen de vereiste randvoorwaarden en die breed toepasbaar, doorzichtig en duurzaam is. Met een voorbeeld wordt het probleemgebied geanalyseerd en de oplossing stapsgewijs beschreven.

Als voorbeeld wordt een winkelketen in bouw- en tuinmaterialen genomen. Deze keten verkoopt een breed assortiment artikelen, van eenvoudige handgereedschappen als snoeischaren en schoffels tot kostbare elektrische apparaten als grasmaaiers, strimmers en schuurmachines. Deze artikelen worden zowel in de eigen winkels verkocht als via franchisenemers.

Jaarlijks wordt op het hoofdkantoor een omzetdoelstelling opgesteld voor de verschillende artikelen. Voor de kostbare artikelen wordt per artikel een omzetdoelstelling opgesteld en voor de goedkopere artikelen per artikelgroep. Daarnaast wordt onderscheid gemaakt tussen de eigen winkels en de franchisenemers. Voor iedere eigen winkel wordt een omzetdoelstelling opgesteld, terwijl dit voor de franchisenemers per regio gebeurt. Een voorbeeld is weergegeven in afbeelding 1.

Het bedrijf heeft een datawarehouse gerealiseerd waarin, naast andere bedrijfsgegevens, de omzetdoelstellingen en de verkooptransacties worden vastgelegd. Het gegevensmodel van

het datawarehouse is gebaseerd op sterschema-modellering. Het wisselende detailniveau van de omzetdoelstellingen stelt eisen aan het gegevensmodel van het datawarehouse en aan de wijze waarop gerapporteerd kan worden.

Uitwerking

In het hierboven geschetste probleem spelen de begrippen *omzetdoelstelling*, *verkooptransactie*, *artikel* en *verkooppunt* de belangrijkste rollen.

De omzetdoelstelling en de verkooptransactie zijn de *meetwaarden of feiten* waar het in dit voorbeeld om draait. Het artikel en het verkooppunt zijn de *invalshoeken of dimensies* waarlangs deze meetwaarden aan elkaar gerelateerd en met elkaar vergeleken worden.

Verkooptransacties zijn afkomstig uit de transactieverwerkende systemen van het bedrijf en hebben een grote mate van detail. Een verkooptransactie bevat elementaire informatie als datum en tijdstip van de transactie, het nummer van de kassa waar de transactie heeft plaatsgevonden, het artikelnummer van het verkochte artikel en het verkochte aantal.

De verzameling verkooptransacties wordt in dit artikel aangeduid als de *realisatiecijfers*. De omzetdoelstellingen worden aangeduid als *plancijfers*.

Over het algemeen worden plancijfers niet opgesteld op het detailniveau van de realisatie.

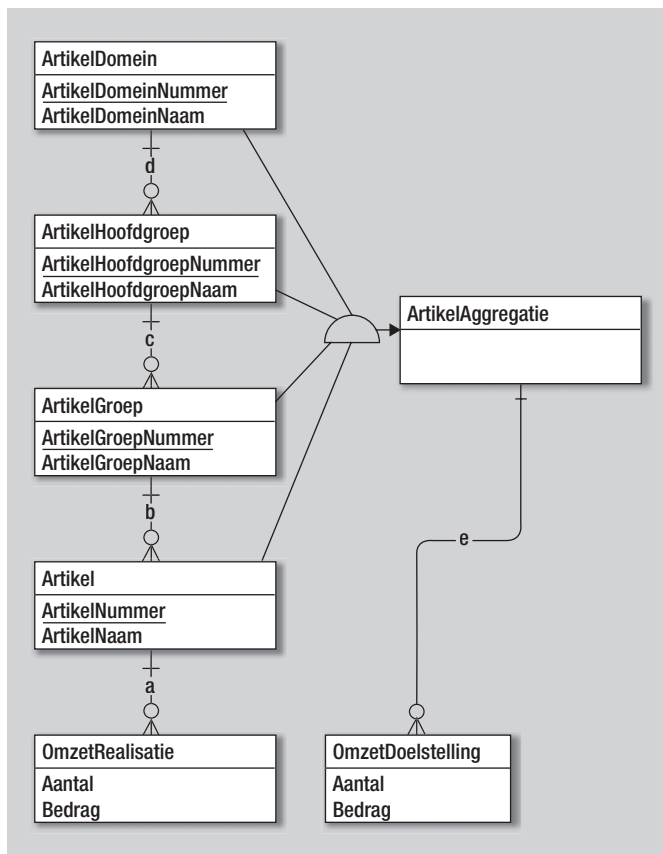
Meestal worden alle plancijfers op één aggregatieniveau opgesteld, bijvoorbeeld per artikelgroep en per regio. In dit geval echter varieert het aggregatieniveau van de plancijfers per artikel en per verkooppunt. Hetzelfde kan zich ook voordoen in bijvoorbeeld de dimensie tijd. De gegevensstructuur van het datawarehouse moet deze flexibiliteit ondersteunen. Hoe kan het datamodel hierin voorzien?

Uitwerking voor de dimensie 'Product'

De gekozen oplossingsrichting is gebaseerd op een algemeen principe, dat voor alle dimensies op een zelfde manier kan worden toegepast. Dit principe kan duidelijk gemaakt worden aan de hand van de product-dimensie. Binnen deze dimensie wordt met een 'artikel' het tastbare, verhandelbare product aangeduid. Deze artikelen worden ten bate van besturing en planning gegroepeerd naar artikelgroepen. Deze op hun beurt worden

Winkel/Regio	Artikel/Artikelgroep	Plan (€)
Zoetermeer	Grasmaaier	1000
Zoetermeer	Strimmer	350
Zoetermeer	Tuingereedschap niet-elektrisch	200
Zoetermeer	Sanitair	3000
Leiden	Grasmaaier	800
Leiden	Strimmer	400
Leiden	Tuingereedschap niet-elektrisch	250
Franchise ZH	Tuingereedschap niet-elektrisch	750

Afbeelding 1: Omzetdoelstellingen.



Afbeelding 2: Gegevensmodel van product-hiërarchie, omzetrealisatie en omzetdoelstelling.

gegroepeerd in artikelhoofdgroepen, en de artikelhoofdgroepen worden tenslotte gegroepeerd in artikeldomeinen. Wanneer de gegevensanalyse voor de product-dimensie wordt vertaald naar een gegevensmodel, dan kan dit er uit zien zoals in afbeelding 2; hierin zijn ook de omzetrealisatie en de omzetdoelstelling opgenomen.

In het model in afbeelding 2 is de realisatie gekoppeld aan het artikel-niveau. De doelstelling is niet gekoppeld aan een specifiek niveau, maar aan een generalisatie van de objecten in de product-hiërarchie. Deze generalisatie is 'Artikelaggregatie' genoemd. In dit model geldt de restrictie dat elk object van het supertype 'Artikelaggregatie' te allen tijde een object van precies één van de subtypen representeert.

Het inbrengen van de realisatiecijfers in het datawarehouse levert geen bijzondere situatie op. De structuur van de plancijfers levert ons echter de volgende uitdagingen:

- Hoe kan deze structuur gerepresenteerd worden in een sterschema;
- Hoe kan deze structuur gevuld worden met data;
- Hoe kan hiermee gerapporteerd worden;
- Hoe vindt de confrontatie met de realisatiecijfers plaats.

Modelleren van het sterschema

In een datawarehouse-ontwerp met feiten- en dimensietabellen vertegenwoordigen de records in de dimensietabellen de

basisobjecten 'onderaan de hiërarchie' (op het detailniveau van de feiten). In de beschreven situatie zal een plancijfer-feit niet noodzakelijk gekoppeld zijn aan het onderste niveau. Het ene feit in de plancijfer-feitentabel zal gekoppeld zijn aan een object op artikel-niveau, terwijl een ander feit in diezelfde feitentabel gekoppeld is aan een object op artikelgroep-niveau.

Een dimensietabel die alleen de objecten op het onderste niveau vertegenwoordigt, voldoet niet in dit geval.

Een mogelijke oplossingsrichting is: 'Verdeel de feiten over verschillende feitentabellen op basis van het niveau in de hiërarchie, creëer aggregaatdimensies voor elk benodigd niveau, koppel de feitentabellen aan de juiste dimensietabellen.'

In afbeelding 5 zijn de feitentabellen als bollen en de dimensietabellen als rechthoeken weergegeven.

Deze oplossing heeft als nadeel dat er een zeer groot aantal feitentabellen en dimensietabellen ontstaat. Zowel voor de beheersbaarheid (uit het oogpunt van beheer en onderhoud) als voor de begrijpelijkheid (uit het oogpunt van de gebruiker) is dit geen goede oplossing.

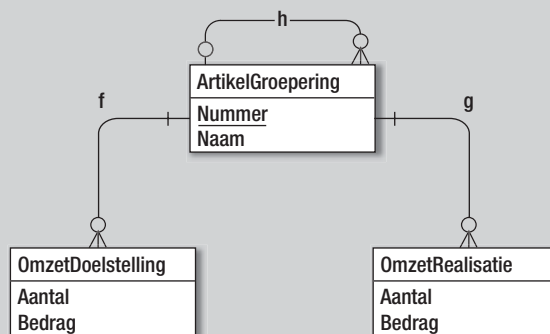
De oplossing: multi-aggregaat dimensie

Om eenvoud in het model, begrijpelijkheid voor de gebruiker en beheersbaarheid voor de laadprocessen te bewerkstelligen, is een andere oplossing ontwikkeld: Hou de feiten in één feitentabel, creëer *multi-aggregaat* dimensies en koppel elk feit aan het dimensie-record van het juiste aggregatieniveau.

In een multi-aggregaat dimensie worden de basisdimensie en alle relevante aggregaat-dimensies samengebracht. Elke groep

Alternatief gegevensmodel

Op basis van de gegevensanalyse kunnen ook verschillende modellen worden opgesteld die voldoen aan de specifieke informatie-behoefte. Eén van de alternatieven is om de verschillende niveaus in de productdimensie te generaliseren tot één object met een zelfverwijzende hiërarchie. Dit is weergegeven in afbeelding 3.

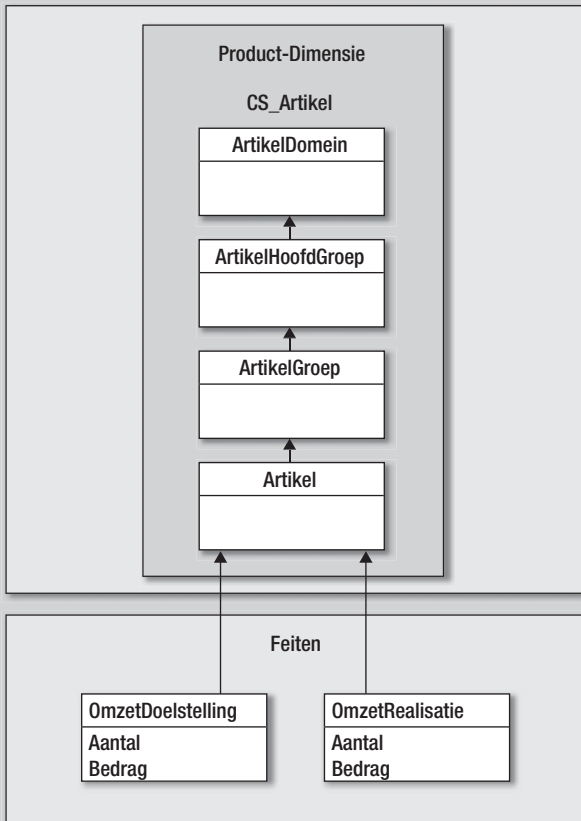


Afbeelding 3: Alternatieve modelleringsoplossing.

Deze oplossingsrichting zal in dit artikel niet verder worden uitgewerkt.

Kalido

In het datawarehouse-product Kalido is voor het beschreven fenomeen een speciale datamodel-constructie beschikbaar, de zogenaamde 'Coding Structure'. Een Coding Structure is een entiteit die een aantal entiteiten in een hiërarchie omvat. Het feit wordt in het model aan deze omvattende entiteit gekoppeld. De feit-records kunnen aan één van de niveaus binnen de Coding Structure worden gekoppeld. Een voorbeeld Coding Structure voor de artikelhiërarchie is weergegeven in afbeelding 4.



Afbeelding 4: Kalido Coding Structure (CS_Artikel).

Een beschrijving van hoe Kalido de gegevens volgens deze structuren vastlegt en hoe deze vervolgens gerapporteerd worden, valt buiten het bestek van dit artikel. De uitgewerkte oplossing in dit artikel komt op onderdelen overeen met de oplossing van Kalido.

dimensie-records vertegenwoordigt een specifiek aggregatieniveau. De product-dimensietabel bevat in dit geval dimensie-records voor artikelen, artikelgroepen, artikelhoofdgroepen en artikeldomeinen. De tabel in afbeelding 6 geeft een gedeelte van de inhoud van deze product-dimensietabel weer. De verschillende groepen zijn in een lichtere tint aangegeven.

Een plancijfer dat een specifieke doelstelling voor grasmaaiers bevat, wordt dan als feit gekoppeld aan het dimensie-record met Artikel-key = 1, terwijl een plancijfer voor alle niet-elektrisch

tuingereedschap als feit gekoppeld wordt aan het dimensie-record met Artikel-key = 10. De realisatiecijfers zullen altijd aan dimensie-records op artikelniveau gekoppeld worden.

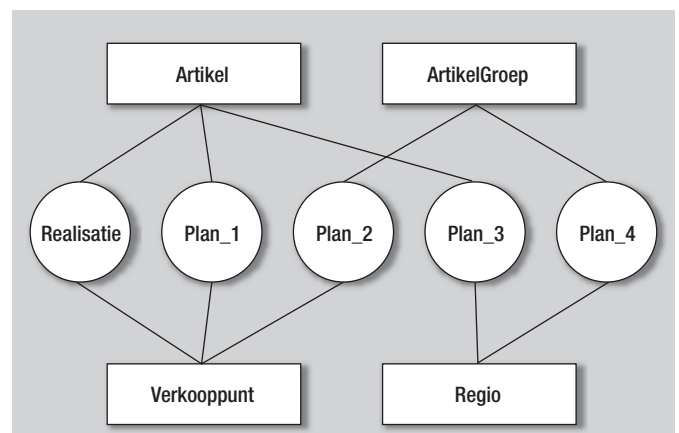
Het daadwerkelijke koppelen van de feiten aan deze dimensie-records levert nog problemen op. Het zoeken naar een dimensie-record met artikelgroep-nummer 56 levert drie records op. Alleen zoeken op artikelgroep-nummer is dus niet voldoende. We moeten ook nog aangeven dat we een dimensie-record nodig hebben van het juiste aggregatieniveau. Deze informatie moet daarom worden toegevoegd aan de dimensie-records.

Dit aggregatieniveau komt overeen met de objectklasse (of entiteit) in het corresponderende genormaliseerde datamodel. Daarom wordt de term *objectklasse* gebruikt als attribuutnaam in het dimensie-record. De term 'niveau' is in dit geval niet correct, omdat het suggereert dat alle entiteiten in een dimensie in één rangorde kunnen worden geplaatst. Dit is in vele modellen niet het geval. Te denken valt aan alternatieve producthiërarchieën.

Naast het attribuut objectklasse wordt ook nog een extra attribuut *objectid* aan het dimensie-record toegevoegd. Dit attribuut bevat de identificatie (logische sleutel) van het object waar dit dimensie-record naar verwijst. Dit is weliswaar redundant, want de sleutels zijn ook als reguliere attributen aanwezig, maar zal toch zijn nut bewijzen aan de rapportagekant. De dimensietabel ziet er dan uit zoals in afbeelding 7.

Het koppelen van feiten aan deze dimensie is nu ondubbelzinnig mogelijk. Bij het plancijfer voor ArtikelGroep 56 wordt als volgt de juiste dimensie-record gevonden: Zoek dimensie-record waarvoor ObjectKlasse = "ArtikelGroep" en ObjectID = "56".

Naast het laden van de feiten in de feitentabel moeten ook de dimensiegegevens geladen worden. De meest eenvoudige benadering hiervoor is door de dimensietabel te 'partitioneren' met het attribuut 'objectklasse' als partitionerings-sleutel en vervolgens elke partitie te behandelen als een separate dimensietabel. In de praktijk zal met de meeste ETL-oplossingen een elegantere aanpak mogelijk zijn. Het betreffende onderdeel van het datawarehouse-model is te zien in afbeelding 8.



Afbeelding 5: Oplossing met veelheid aan feitentabellen.

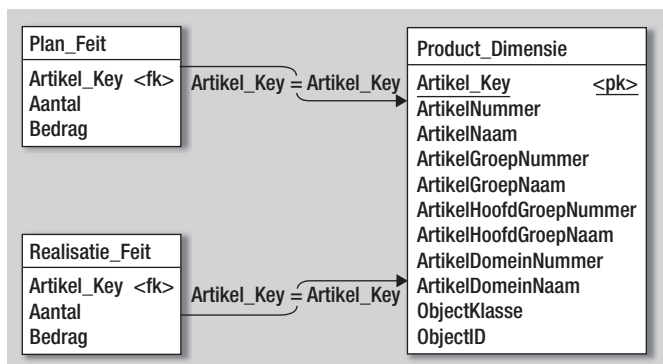
Artikel-Key	Artikel	Artikel-Naam	Artikel-Groep	Groep-Naam	Artikel-HoofdGroep	HoofdGroep-Naam	Artikel-Domein	Domein-Naam
1	025900	Grasmaaier	55	Elektrisch	10	Tuingereedschap	2	Tuin
2	024224	Strimmer	55	Elektrisch	10	Tuingereedschap	2	Tuin
3	342231	Hark	56	Niet-elektrisch	10	Tuingereedschap	2	Tuin
4	342232	Schoffel	56	Niet-elektrisch	10	Tuingereedschap	2	Tuin
9			55	Elektrisch	10	Tuingereedschap	2	Tuin
10			56	Niet-elektrisch	10	Tuingereedschap	2	Tuin
21					10	Tuingereedschap	2	Tuin
33							2	Tuin
34							3	Sanitair

Afbeelding 6: Inhoud van de multi-aggregaat dimensietabel.

Artikel-Key	Artikel	Artikel-Naam	Artikel-Groep	Groep-Naam	Artikel-HoofdGroep	HoofdGroep-Naam	Artikel-Domein	Domein-Naam	Object-Klasse	Object-ID
1	025900	Grasmaaier	55	Elektrisch	10	Tuingereedschap	2	Tuin	Artikel	025900
2	025901	Strimmer	55	Elektrisch	10	Tuingereedschap	2	Tuin	Artikel	025901
3	342231	Hark	56	Niet-elektrisch	10	Tuingereedschap	2	Tuin	Artikel	342231
4	342232	Schoffel	56	Niet-elektrisch	10	Tuingereedschap	2	Tuin	Artikel	342232
9			55	Elektrisch	10	Tuingereedschap	2	Tuin	ArtikelGroep	55
10			56	Niet-elektrisch	10	Tuingereedschap	2	Tuin	ArtikelGroep	56
21					10	Tuingereedschap	2	Tuin	ArtikelHoofdGroep	10
33							2	Tuin	ArtikelDomein	2
34							3	Sanitair	ArtikelDomein	3

Afbeelding 7: Dimensietabel inclusief objectklasse en objectid.

De beschreven oplossing is generiek. Voor elke dimensie in het sterschema kan deze oplossing op dezelfde wijze worden toegepast. De toepassing van de oplossing stelt ook voorwaarden. Zo is het nodig dat alle primaire sleutels van de verschillende objectklassen naar één uniform datatype afgebeeld kunnen worden. Dit is nodig voor de vulling van het objectid. Daarnaast dienen de namen van de objectklassen beheerd te worden. Metadata worden in deze oplossing ook als data gebruikt.



Afbeelding 8: Stermodel (vereenvoudigd) met twee feitentabellen en een gemeenschappelijke dimensietabel.

Rapportage

De volgende uitdaging met betrekking tot deze plancijfers is het produceren van rapportages. Het variabele niveau waarop de feiten beschikbaar zijn maakt een rapportage anders dan wanneer alle feiten op hetzelfde niveau zijn ingevoerd. Als voorbeeld worden de plancijfers voor de winkel in Zoetermeer gebruikt, zoals aan het begin beschreven. Deze plancijfers worden in de feitentabel vastgelegd zoals te zien in afbeelding 9. Door in de rapportage bij ontbrekende waarden op bepaalde niveaus een standaardwaarde (default) te gebruiken, kunnen alle gewenste rapportages geproduceerd worden. Een rapportage van de omzetdoelstellingen gegroepeerd op artikelniveau ziet er dan uit als in afbeelding 10.

Artikel_Key	OmzetDoelstelling
1	1000
2	350
10	200
34	3000

Afbeelding 9: Feitentabel met plancijfers.

Artikel	Plan (€)
Grasmaaier	1000
Strimmer	350
(blank)	3200
Totaal	4550

Afbeelding 10.

Artikelgroep	Plan (€)
Elektrisch	1350
Niet-Elektrisch	200
(blank)	3000
Totaal	4550

Afbeelding 11.

En een rapportage op het hogere artikelgroep-niveau ziet er uit als in afbeelding 11. Veel gebruikte rapportage- en analyse-tools hebben functionaliteiten om met deze 'missing values' om te gaan, zowel in de traditionele lijst-georiënteerde rapportages als in OLAP-kubussen. Wanneer er nu bijvoorbeeld zowel voor de hele artikelgroep 'Niet-Elektrisch' als voor het specifieke artikel 'Schoffel' een omzetdoelstelling is vastgesteld is het van belang te weten of de groepsdoelstelling inclusief of exclusief de specifieke doelstelling is. Dit bepaalt namelijk de wijze van berekening in bovenstaande rapportage.

Een andere, voor de hand liggende rapportage is een lijst van de individuele plancijfers zoals ze zijn ingevoerd. Deze rapportage moet dan verschillende niveaus door elkaar heen presenteren. De dimensietabel biedt hiervoor in ieder geval de niveau-onafhankelijke velden 'objectklasse' en 'objectid'. Hiermee kan de rapportage gemaakt worden, zonder gebruik te moeten maken van conditionele expressies in de rapportagetool, zie afbeelding 12. Wat hier ontbreekt is een zinvolle omschrijving. Om deze toe te kunnen voegen aan de rapportage kunnen twee scenario's gevolgd worden:

1. Via conditionele expressies in het rapportagescript wordt de juiste omschrijving bepaald;
2. Aan de dimensietabel wordt naast de velden 'ObjectKlasse' en 'ObjectID' ook nog een generiek veld 'ObjectNaam' toegevoegd, dat kan worden gebruikt in de rapportage.

Groepering	Code	Plan (€)
Artikel	025900	1000
Artikel	025901	350
ArtikelGroep	56	200
ArtikelDomein	3	3000
Totaal		4550

Afbeelding 12.

Beide scenario's zijn mogelijk, een onderbouwing van de keus wordt hier niet verder uitgewerkt. De relevantie van de totaal-telling in deze rapportage hangt af van de volledigheid van de plancijfers. Wanneer er twee plancijfers op verschillend niveau in dezelfde hiërarchische lijn zijn vastgesteld, is het totaal in ieder geval niet relevant, omdat hierin namelijk een dubbel telling plaatsvindt.

De volgende stap is de confrontatie tussen plan- en realisatiecijfers. Bij deze confrontatie moeten feiten op verschillend aggregatieniveau aan elkaar worden gerelateerd.

De volgende verkooptransacties hebben plaatsgevonden: een grasmaaier verkocht voor € 200; een strimmer verkocht voor € 150; een grasmaaier verkocht voor € 340; een hark verkocht voor € 25; een strimmer verkocht voor € 150; een hark verkocht voor € 25; een hark verkocht voor € 15. Dit levert de feitentabel op, zoals in afbeelding 13.

Artikel_Key	Omzet
1	200
2	150
1	340
3	25
2	150
3	25
3	15

Afbeelding 13: Feitentabel met realisatiecijfers.

De confrontatie tussen de realisatiecijfers en de plancijfers kan op verschillende manieren tot stand komen: beide cijfers worden op een specifiek niveau geaggregeerd en vervolgens aan elkaar gerelateerd; de realisatiecijfers worden geaggregeerd naar de individuele ingevoerde plancijfers.

Het eerste scenario is een 'standaard scenario' voor de rapportage- en analyse-tools. Hier schuilt echter een addertje onder het gras. Als bijvoorbeeld een rapportage op artikelniveau wordt gemaakt, dan zal deze er voor de realisatie uitzien als in afbeelding 14. De cijfers voor de grasmaaier en de strimmer hebben een equivalent aan de plancijfer-kant, maar er bestaat geen plancijfer voor de hark. Deze maakt echter wel deel uit van het plancijfer dat aan '(blank)' is gekoppeld. Een goede confrontatie is in dit geval niet mogelijk. Dit eerste scenario kan alleen worden gebruikt vanaf het

ArtikelCode	Omschrijving	Realisatie (€)
025900	Grasmaaier	540
025901	Strimmer	300
342231	Hark	65
	Totaal	905

Afbeelding 14.

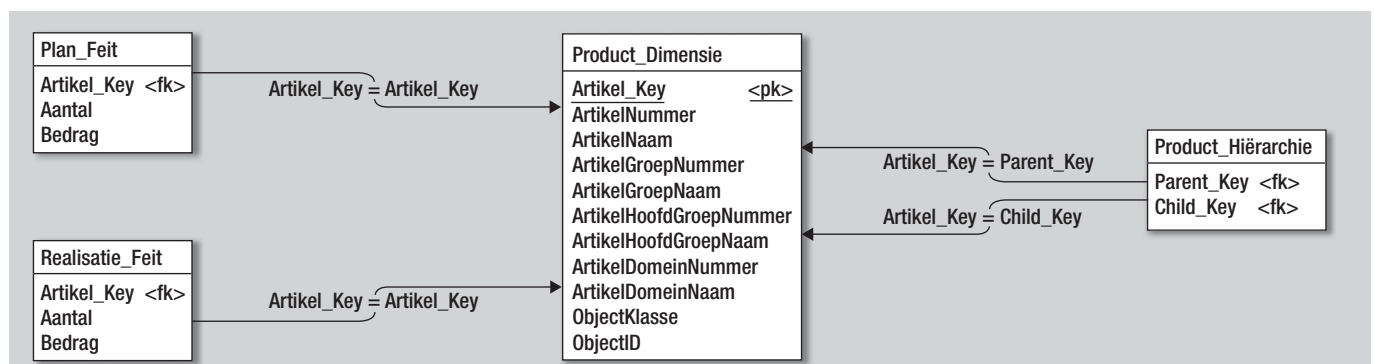
aggregatieniveau waar geen (blank) meer voorkomt. Het tweede scenario levert met de gegevens van het gebruikte voorbeeld de rapportage op, te zien in afbeelding 15. Met andere woorden: bij elk record in de plan-feitentabel worden de corresponderende realisatiecijfers gezocht. Ook hier geldt weer dat de totaalstelling niet altijd zinvol is. Deze telling is dan zinvol wanneer het een relevante groepering vertegenwoordigt, én het berekende bedrag voor deze groepering juist is (en dus geen dubbelstellingen bevat).

Het gebruik van een brugtabel

Om de realisatiecijfers op een juiste manier aan de plancijfers te koppelen kan niet worden volstaan met het koppelen van de dimensie-sleutels van beide feitentabellen. De koppeling tussen de twee feitentabellen kan op een juiste manier worden gelegd met behulp van een brugtabel. Deze brugtabel is een hulptabel behorende bij een specifieke dimensie. In deze brugtabel worden de relaties tussen de verschillende dimensie-records vastgelegd. In de brugtabel worden alle 'Parent-Child' relaties vastgelegd, waarbij geldt dat een object ook zijn eigen nakomeling is. Voor de hierboven beschreven product-dimensietabel bevat de brugtabel in totaal 26 records. Het datamodel waarin de brugtabel is opgenomen is te zien in afbeelding 16. Met behulp van de brugtabel kan de bedoelde rapportage eenvoudig gemaakt worden. De SQL voor het aggregeren van de realisatiecijfers per ingevoerd plancijfer is weergegeven in afbeelding 17. De MAX- en SUM-functies zorgen ervoor dat de juiste bedragen worden berekend. Het resultaat van deze SQL query kan vervolgens gekoppeld worden aan de dimensietabel om de gevraagde rapportage op te leveren. Het onderhoud van de brugtabel vindt plaats nadat de dimensietabel is bijgewerkt. De meest eenvoudige methode

Niveau	Code	Omschrijving	Plan (€)	Realisatie (€)
Artikel	025900	Grasmaaier	1000	540
Artikel	025901	Strimmer	350	300
ArtikelGroep	56	Tuingereedschap		
		niet-elektrisch	200	65
ArtikelDomein	3	Sanitair	3000	0
Totaal			4550	905

Afbeelding 15: Rapportage op basis van het tweede scenario.



Afbeelding 16: Datamodel inclusief brugtabel.

```

SELECT p.Artikel_key, MAX(p.OmzetDoelstelling), SUM(r.Omzet)
FROM Plan_Feit p, Realisatie_Feit r, Artikel_Hiërarchie h
WHERE p.Artikel_Key = h.Parent_Key
AND r.Artikel_Key = h.Child_Key
GROUP BY p.Artikel_Key
  
```

Afbeelding 17: SQL voor het aggregeren van realisatie per plancijfer.

```

INSERT INTO Product_Hiërarchie (Child_Key, Parent_Key)
SELECT s.Artikel_Key as Child_Key, p.Artikel_Key as Parent_Key
FROM Product_dimensie s, Product_dimensie p
WHERE s.ArtikelGroepNr = p.ArtikelGroepNr
AND p.ObjectKlasse = 'ArtikelGroep';
  
```

Afbeelding 18: SQL voor het vullen van de brugtabel.

hiervoor is het steeds opnieuw aanmaken van alle records in de tabel. Per objectklasse wordt de dimensietabel aan zichzelf gekoppeld (inner self join) met als clause dat de betreffende sleutel (niet het objectid!) aan beide zijden gelijk moet zijn. Elk resultaat-record genereert een record in de brugtabel. Voor de objectklasse 'ArtikelGroep' is de SQL zoals weergegeven in afbeelding 18.

Tot besluit

Het gebruik van 'multi-aggregaat'-dimensietabellen is een generieke oplossing voor het modelleren van een sterschema als feiten een variabel detailniveau in de dimensie-hiërarchieën kennen. De oplossing is toepasbaar voor alle dimensies waar deze variabiliteit zich voordoet. De brugtabel helpt bij het koppelen van de hiërarchische niveaus.

Het toepassingsgebied van deze oplossing reikt verder dan alleen het probleem dat in dit artikel is beschreven. De 'multi-aggregaat'-dimensietabellen kunnen bijvoorbeeld ook gebruikt worden voor algemene aggregatie-doeleinden. In dit artikel is dit buiten beschouwing gelaten. Ook is in de uitwerking van de oplossing het fenomeen 'historie' buiten beschouwing gelaten. Het bijhouden van historie (slowly changing dimensions) betekent een toename van de complexiteit in het onderhoud van de brugtabel.

Mark Zwijsen

Drs. M.C.M. Zwijsen (mark.zwijsen@atosorigin.com) is Senior Consultant Data Warehousing bij Atos Origin.