

Minidimensies en rapportages in de praktijk

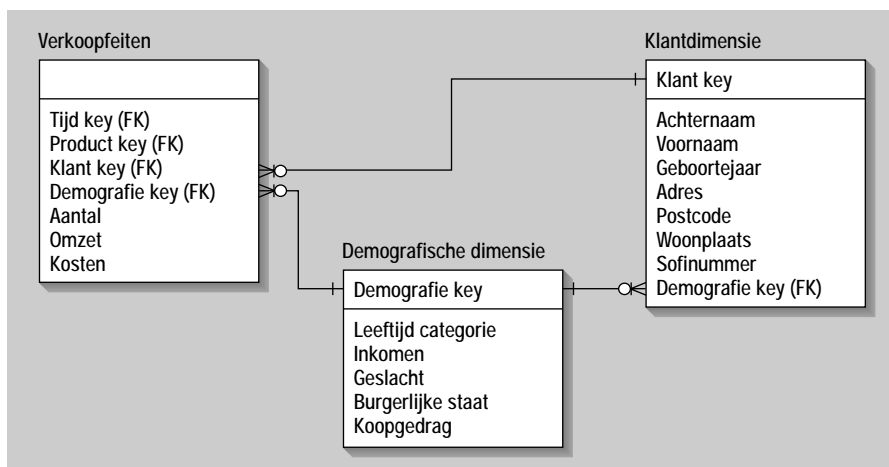
‘Kimball-sterren’ voor flexibele informatie

Alexander van Helm en Murad Maas

Dit artikel beschrijft twee praktijkvoorbeelden die laten zien hoe zogenaamde Kimball-sterren gebruikt kunnen worden voor flexibele informatievoorziening. Het eerste voorbeeld toont hoe het historisch effect van minidimensies is te creëren met behulp van procesdata. Uit het tweede voorbeeld blijkt hoe specifieke rapportagegroepen makkelijk geselecteerd kunnen worden.

Kimball-sterren worden vooral gebruikt voor datamarts. Dimensies moeten het verhaal vertellen over de gegevens. In de meeste ontwerpen van dimensies bestaan daardoor -op de technische datawarehousesleutel en -geldigheidsperiode na- alleen velden die voor de informatievoorziening interessant zijn. De ETL-programmatuur die de ster vult bevat de afleidingsregels, die aangeven welk dimensierecord bij een bepaald feit hoort.

De afleidingsregels zijn de business-rules: de logica die vertelt hoe een organisatie naar haar gegevens kijkt. Zijn de business-rules ingebouwd in de ETL-programmatuur, dan zorgt een wijziging daar-



FIGUUR 1: KLASSIEK VOORBEELD VAN AFGESPLITSTE ATTRIBUTEN TOT MINIDIMENSIE.

van voor verandering in de ETL-programmatuur. Maar ook zullen miljoenen records moeten worden aangepast, en dat is zonde van de ICT-moeite. Kunnen we dat voorkomen?

GEBRUIK VAN PROCESDATA

Kan het datawarehouse ervoor zorgen dat de actuele afleiding beschikbaar is? Zeker kan dat. Door dimensies uit te brei-

den of zelfs puur te creëren met die procesgegevens, die bepalend zijn voor de afleiding van de actuele situatie. Tevens door de attributen waarvan de actuele situatie interessant is, af te splitsen als minidimensie. Dit is mogelijk dankzij het historisch effect van die dimensies.

Een minidimensie is een stel attributen, die afgesplitst worden uit een dimensie om op hun beurt een nieuwe dimensie te vormen. De oude dimensie verwijst met behulp van de datawarehousesleutel naar de minidimensie. In de oude dimensie wordt de datawarehousesleutel bijgewerkt (historie type 3) naar de actuele situatie. De historische verwijzing is te vinden in de feitentabel. Meestal wordt een minidimensie afgesplitst om teveel groei in de dimensie (historie type 2) te voorkomen.

We geven een klassiek voorbeeld, uit de cursus van Harm van der Lek. Door met name de attributen leeftijdscategorie

Definitie Kimball-ster

Een Kimball-ster is een datamodel speciaal ontworpen voor informatievoorziening. Hij bestaat uit een feitentabel met daaraan gekoppeld dimensies. Feittabellen bevatten de meetwaarden, de te rapporteren gegevens. Dimensies vormen de business-'kapstok'. Ze omschrijven hoe men naar gegevens moet kijken. Denk hierbij aan hierarchieën en dergelijke. Ze geven aan hoe de gegevens aan elkaar relateren. Het bijzondere aan een Kimball-ster is dat bij een join altijd een rechtstreekse koppeling bestaat tussen de feitentabel en de betrokken dimensie.

en inkomen af te splitsen van de klantdimensie vermijden we veelvuldig stapelen; zie figuur 1.

MINIDIMENSIES EN HISTORIE

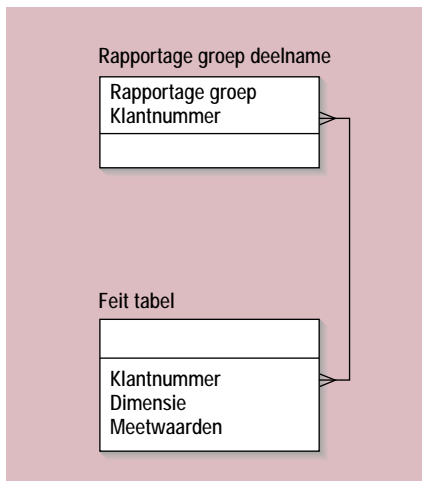
Wat is het effect van minidimensies op de historie? In de oude (klant)dimensie staat de actuele verwijzing naar de minidimensie. Nu kunnen we door een rechtstreekse koppeling van de feitentabel naar de minidimensie kijken naar de demografische waarden ten tijde van het feit en door een koppeling naar de minidimensie via de klantdimensie naar de actuele demografische waarden.

Hoewel dit in de literatuur als bijkomend effect wordt gepresenteerd, kun je in de praktijk dit effect juist opzoeken. We proberen dat toe te lichten aan de hand van een praktijkvoorbeeld.

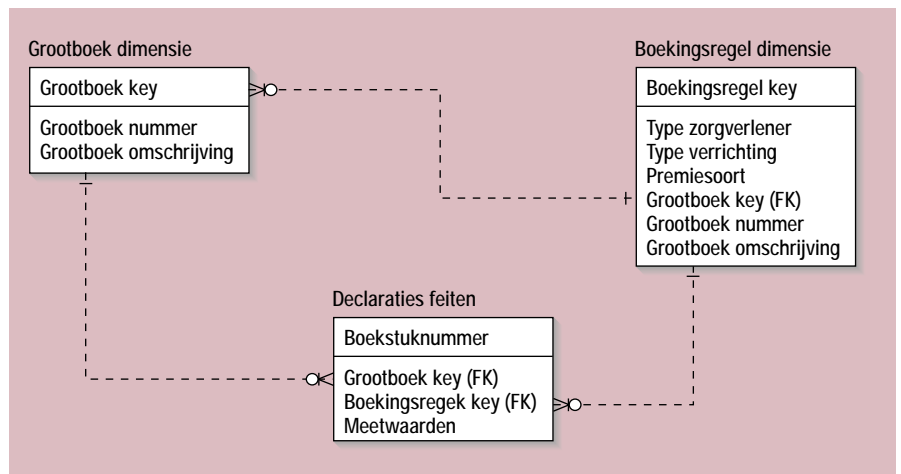
Bij de Agis Groep -en waarschijnlijk ook bij andere zorgverzekeraars- wordt de grootboekrekening waarop een declaratie drukt, bepaald door een aantal zaken:

- type zorgverlener (zorgverlenersoort en specialisme);
- type verrichting (honoreringsoort en honoreringcode);
- premiesoort (welk pakket dekt de declaratie).

Dit zijn procesgegevens, die we kunnen omschrijven als boekingsregel. De boekingsregels kunnen per jaar veranderen. In het ene jaar kan een zelfde soort declaratie op een andere grootboekrekening



FIGUUR 3.



FIGUUR 2: GROOTBOEKDIMENSIE-ATTRIBUTEN OPGENOMEN IN DE BOEKINGSREGELDIMENSIE.

drukken dan in het andere jaar. Voor het bedrijf is het interessant te weten op welke grootboekrekening een schade drukt bij registratie, maar ook waarop de declaraties nu zouden drukken. Met dat in het achterhoofd is de volgende oplossing bedacht.

In plaats van alleen een grootboekdimensie te creëren is een boekingsregeldimensie aangemaakt. Deze dimensie heeft als attributen de procesgegevens van het grootboek (type zorgverlener, type verrich-

Het datawarehouse kan ervoor zorgen dat de actuele afleiding beschikbaar is

ting en premiesoort) en alle attributen van de grootboekdimensie. Deze is een minidimensie van de boekingsregeldimensie.

Doordat de boekingsregel tijdens de verwerking wordt bijgewerkt naar de actuele situatie zijn declaraties historisch (grootboekdimensie) en actueel (boekingsregeldimensie) te bekijken. Voor de helderheid van het model zijn alle grootboekdimensie-attributen opgenomen in de boekingsregeldimensie. Dit betekent wel dat men elke verandering in de grootboekdimensie ook moet doorvoeren in de boekingsregeldimensie (zie figuur 2).

We werken het voorbeeld uit met fictieve gegevens. Stel dat huisartsbezoeken eerst drukken op grootboekrekening 'huisarts algemeen' (5) en vanaf een bepaald moment op 'huisarts praktijk' (6). De enige

wijziging in de data is dan het verwijzende record in de actuele grootboekdimensie. Kijk wat er gebeurt met volgnummer 21 van boekstuknummer 9812000001 in tabel 1, met betrekking tot drie huisartsbezoeken voor 120 gulden, in de tabellen 2, 3 en 4.

Een join van declaratiefeiten met de twee dimensies, levert vóór de verandering het resultaat op van tabel 5. En na de verandering ontstaat tabel 6.

Zo kunnen we een ingewikkeld probleem oplossen, door een dimensie te creëren met procesgegevens die bepalend zijn voor de actuele situatie (type zorgverlener, type verrichting en premiesoort) en de attributen waarvan zowel de actuele als historische waarde interessant is (de grootboekdimensie) af te splitsen als minidimensie. Natuurlijk moet de ETL-programmatuur nog wel ontwikkeld worden. Vooral voor prognoses en 'what if'-analyses is dit gebruik van minidimensies zeer interessant.

RESULTAAT

De procesgegevens hoeven voor de gebruiker niet zichtbaar te zijn. Datamarts worden hetzij via een logische laag bekeken of doorgeladen in een OLAP-tool; in de praktijk bij Agis is dat Business Objects. De procesgegevens worden dan niet aan de gebruiker getoond. Deze heeft beschikking over een actueel grootboek, een historisch grootboek of in sommige gevallen zelfs beide.

Het gewenste resultaat is bereikt. Deze werkwijze is natuurlijk op meer problemen

Declaratie feiten

Boekstuk-nummer	Volg-nummer	Actueel grb	Historisch grb	Aantal	Bedrag
...
9812000001	21	5	5	3	120
...

TABEL 1.

Grootboek dimensie

Key	Omschrijving
...	...
5	Huisarts algemeen
6	Huisarts praktijk
...	...

TABEL 2.

Boekingsregel dimensie voor verandering

Key	Zorg-verlener	Honorerings-code	Premie-soort	Grb key	Omschrijving
...
123	1010	1234567890	01011010	5	Huisarts algemeen
...

TABEL 3.

Boekingsregel dimensie na verandering

Key	Zorg-verlener	Honorerings-code	Premie-soort	Grb key	Omschrijving
...
123	1010	1234567890	01011010	6	Huisarts praktijk
...

TABEL 4.

Boekstuk-nummer	Volg-nummer	Actueel grb	Historisch grb	Aantal	Bedrag
...
9812000001	21	Huisarts algemeen	Huisarts algemeen	3	120
...

TABEL 5.

Boekstuk-nummer	Volg-nummer	Actueel grb	Historisch grb	Aantal	Bedrag
...
9812000001	21	Huisarts praktijk	Huisarts algemeen	3	120
...

TABEL 6.

dan het grootboek alleen toepasbaar, en de techniek is eventueel ook in sneeuwvlok-schema's te gebruiken.

RAPPORTAGEGROEPEN

Vaak wordt met veel moeite een specifieke klantengroep bepaald, waarop allerlei rapportages gemaakt moeten worden. Deze klantengroep is niet eenvoudig te bepalen en bestaat misschien wel uit duizenden klanten. Men moet daarom een mechanis-

Een eenvoudige beperking is tot een selectieve groep te maken

me bedenken dat deze groep bewaart en inzetbaar houdt voor andere rapportages.

In de praktijk wordt veelal gekozen voor de aanmaak van tijdelijke tabellen met daarin de klantnummers van deze groep. Dit werkt prima, zolang het data-warehouse maar door een zeer selecte groep mensen wordt bekeken, die zelf ook werken aan de ontwikkeling daarvan. Zodra het warehouse echter breder ingezet wordt, zal de organisatie moeten waken voor een wildgroei van tabellen. Ook kost dit erg veel onderhoud op de logische laag: kenbaar maken en joinen van de tijdelijke tabellen.

De Agis Groep maakte veelvuldig gebruik van steekproeven. Denk hierbij aan bepaalde selectiegroepen, bijvoorbeeld suikerpatiënten, en een tegengroep voor medische onderzoeken. Steekproeven zijn namelijk eigenlijk niets meer dan rapportagegroepen, waarvoor je de afleiding niet telkens opnieuw wilt doen. Bij Agis moest een flexibelere methode worden bedacht dan het gebruik van tijdelijke tabellen. Dat is gelukt, en de gevonden oplossing is vrij eenvoudig in een willekeurig datawarehouse te implementeren.

- Maak een brugtabel met daarin klantnummers gekoppeld aan een rapportagegroep.
- Zorg dat in alle feiten het klantnummer als *de-generate dimension* opgeslagen is.
- Een nieuwe rapportagegroep laad je

Boekstuknummer	Volgnummer	Klantnummer	Aantal	Bedrag
9812000001	5	123456789	5	120
9812000001	8	123456789	2	250
9812000001	12	123456789	8	300
9812000013	2	123456789	2	220
9812000001	12	987654321	8	300
9812000003	1	987654321	1	60
9812000003	4	987654321	4	80
9806000001	1	123451234	2	120
Totaal			32	1450

TABEL 7: DECLARATIES.

Klant	Rapportagegroep
123456789	Suiker
123451234	Suiker
987654321	Hart
123451234	Hart

TABEL 8: RAPPORTAGEGROEPEN.

Boekstuknummer	Volgnummer	Klantnummer	Aantal	Bedrag
9812000001	5	123456789	5	120
9812000001	8	123456789	2	250
9812000001	12	123456789	8	300
9812000013	2	123456789	2	220
9806000001	1	123451234	2	120
Totaal			19	1010

TABEL 9: DECLARATIES VOOR RAPPORTAGEGROEP SUIKER.

Boekstuknummer	Volgnummer	Klantnummer	Aantal	Bedrag
9812000001	12	987654321	8	300
9812000003	1	987654321	1	60
9812000003	4	987654321	4	80
9806000001	1	123451234	2	120
Totaal			15	560

TABEL 10: DECLARATIES VOOR RAPPORTAGEGROEP HART.

Boekstuknummer	Volgnummer	Klantnummer	Aantal	Bedrag
9812000001	5	123456789	5	120
9812000001	8	123456789	2	250
9812000001	12	123456789	8	300
9812000013	2	123456789	2	220
9812000001	12	987654321	8	300
9812000003	1	987654321	1	60
9812000003	4	987654321	4	80
9806000001	1	123451234	2	120
9806000001	1	123451234	2	120
Totaal			34	1570

TABEL 11: DECLARATIES VOOR RAPPORTAGEGROEP SUIKER OF HART.

door de tabel uit te breiden met de rapportagegroepgegevens. Hiervoor is geen technische aanpassing nodig.

Zuiver gezien is de rapportagegroepbrug-tabel overigens een *factless fact*-tabel over deelname aan rapportagegroepen.

Het is nu mogelijk feiten voor een rapportagegroep te bekijken door via de brug-tabel deze rapportagegroep als voorwaarde te stellen. Pas wel op dat de voorwaarde

Deze werkwijze is natuurlijk op meer problemen dan het grootboek alleen toepasbaar

altijd op één rapportagegroep of op elkaar uitsluitende rapportagegroepen wordt gelegd, omdat anders dubbelstellingen kunnen ontstaan. Via een logische laag kan dit ook aan de gebruikers beschikbaar gesteld worden. (Zie figuur 3.)

Een voorbeeld uitgewerkt met fictieve data geven de tabellen 7 tot en met 11. Let in de laatste, 'Declaraties voor rapportagegroep suiker of hart', op de dubbelstelling.

ROBUUST EN FLEXIBEL

Uit beide voorbeelden blijkt dat, met een beetje creativiteit, Kimball-sterren robuuste en toch flexibele oplossingen kunnen bieden voor op het eerste gezicht lastige problemen. Zowel een actueel als historisch perspectief is zichtbaar te maken door het creëren van minidimensies. Met de rapportagegroepetechniek kan een eenvoudige beperking tot een selectieve groep worden gemaakt. ●

Alexander van Helm (a.vanhelm@pinkroccade.com) is werkzaam bij PinkRocade en gespecialiseerd in business intelligence & datawarehousing. Hij was en is betrokken bij datawarehouses voor onder meer KLM, Delta Lloyd en Agis Groep. Aan het in dit artikel beschreven project bij Agis heeft ook Van Helms collega Gert-Jan Kooren gewerkt. Murad Maas is werkzaam bij Agis Groep als informatie-analist en systeemontwikkelaar. Hij is vanaf het begin betrokken geweest bij de ontwikkeling van het datawarehouse van zijn bedrijf.