



Drie aspecten van datamodellering nader bekeken

Over boekhouden, verplaatsingen en historie

Karien Verhagen

Het bouwen van een solide datawarehouse-infrastructuur blijft een cruciaal onderdeel van BI. Professionele datamodellen blijven daarin centraal staan. Een datawarehouse is toch nog steeds een knap geconstrueerde gegevens-assemblagemachine.

Er is gelukkig geen standaard oplossing voor BI-modellering, maar wel een reeks boeken en cursussen die inzicht geven en je kunnen helpen. Dit artikel doet suggesties voor drie veelvoorkomende datamodelleringsaspecten.

1. Boekhouden met gegevens;
2. Goederenverplaatsingen en commerciële transacties;
3. Historie oplossen en correcties met terugwerkende kracht.

Boekhouden met gegevens

ICT'ers zijn eigenwijze mensen die op een eilandje dingen bedenken die anderen jaren eerder al hebben uitgevonden. Zo ontdekten ze in de jaren negentig Activity Based Costing als een uitstekende methodiek om inzicht te krijgen in de kostenstructuur van een bedrijf. Boekhouders en controllers hadden dat al zo'n twintig jaar eerder bedacht. ICT'ers hebben daar niets van geleerd en blijven bezig met het uitvinden van het wiel.

Misschien voldoet zelfs boekhoudsoftware wel voor BI

De parallellen in verwante wetenschappen zijn veel voedzamer dan ICT zich realiseert. Zo zou ICT voor veel modelleringsvraagstukken een oplossing kunnen vinden in de aloude praktijk van het boekhouden. Wat is een transactie of feit anders dan een journaalpost? Een journaalpost legt een geconsolideerd feit vast. Dat feit wordt gerelateerd aan kostensoorten en kostenplaatsen. Dat zijn dus selectiekenmerken of dimensies. Boekhoudsoftware heeft vaak ook mogelijkheden voor een derde of vierde dimensie, bijvoorbeeld de projectenstructuur. De journaalposten worden verwerkt in grootboeksystemen. Dat zijn strategische aggregaties van feiten. Ze worden vergeleken met het budget (de normen).

Rapportages uit boekhoudsystemen hebben twee basisvormen: de balans en de verlies- en winstrekening. De balans of de proefbalans noemt ICT 'snapshot-technologie'. Voldoende snapshots achter elkaar geven een aardig beeld van de bedrijfsvoering. De balans geeft het saldo. De verlies- en winstrekening geeft inzicht in het resultaat van de bedrijfsprocessen (een soort Key Performance Indicator dus). Voor geldstromen die onderweg zijn, heeft de boekhoudtheorie de kruisposten verzonnen. Kruisposten zijn nog niet geconsolideerde feiten of verplaatsingstransacties onderweg.

ICT'ers zouden voor het verslaan van feiten en dimensies in rapportages eens goed moeten kijken naar boekhouders. Met de gebruiker op zoek naar een goede KPI zouden ze wel eens kunnen stuiten op het equivalent van een *quick ratio*, *ijzeren voorraad*, *solvabiliteitspercentage* of *current ratio*. Misschien voldoet zelfs boekhoudsoftware wel voor BI.

Goederenverplaatsingen en commerciële transacties

Er zijn feiten die op één bepaald moment plaatsvinden en feiten die enige tijd onderweg zijn. Een commerciële transactie is zo'n feit dat op één bepaald moment in de tijd wordt afgesloten. Van optie of reservering wordt een commerciële transactie een feit op het moment dat de transactie bekrachtigd wordt. Schoolvoorbeeld is een verkooptransactie. Daarnaast zijn er transacties die gerealiseerd worden op de tijdsas, die een bepaalde tijd onderweg zijn. Een goederenverplaatsing, bijvoorbeeld het afleveren van een bestelling, vertrekt van een bepaalde plaats op een bepaald tijdstip en komt na zekere tijd op een andere plaats aan. Vaak verschaft de relatie tussen die twee transacties belangrijke informatie voor managers. In het voorbeeld zou dat het antwoord kunnen zijn op de vraag waar de goederen zijn op een bepaald moment. Federal Express is een van de eerste bedrijven, die heeft

ingezien dat deze informatie voor klanten belangrijk is. De data-modellering voor dit soort vraagstukken zou er uit kunnen zien als in afbeelding 1. In de meeste gevallen is het verstandig de relatie te verbinden via een relatierecord. In de praktijk hebben of krijgen dit soort transacties een *n op m* karakter. Daarmee wordt bedoeld dat er meer (n) bestellingen in één vrachtwagen kunnen staan, maar ook dat één bestelling geleverd kan worden in twee of meer (m) vrachtwagens. Het relatierecord geeft exact het verband aan, te weten welke deel van de bestelling meegaat met welke verplaatsing. Wanneer er in de bedrijfsvoering op dit moment nog geen aanleiding is om hier een *n op m* relatie van te maken, is het toch verstandig hierop te anticiperen.

Airliners kregen in de jaren tachtig te maken met het begrip *code sharing*. Dat zijn meer vluchten in één toestel: de KL673 en de NW4523 zaten bijvoorbeeld in hetzelfde vliegtuig.

Vluchtinformatiesystemen kregen daardoor problemen bij de registratie van vluchten. Een vlucht en een fysiek toestel bleken niet meer één op één te corresponderen. Niet alleen konden meerdere vluchten worden uitgevoerd in één toestel, het was zelfs ook mogelijk een vlucht met twee toestellen uit te voeren. Door de vlucht (dus een verplaatsing van een bepaald punt in de tijd naar een ander punt) en de verkochte vlucht (een commercieel gegeven met een bepaald volume) aan elkaar te relateren via een relatierecord, werd het mogelijk op alle aspecten de gewenste informatie te verschaffen.

Een ander modelleringsprincipe dat de luchtvaart ons als eerste heeft geleerd, is de *absolute verwijziging*. Een vlucht die om 03.00 uur PM vertrekt en om 07.00 uur PM aankomt, heeft er in de regel geen 4 uur over gedaan om van A naar B te komen. Aankomsttijden en vertrektijden van vluchten worden daarom geregistreerd in UTC. UTC (Universal Time Control) is een absolute tijdmaat die overal ter wereld aan hetzelfde moment refereert. De lokale tijd is natuurlijk van belang voor de communicatie naar passagiers, maar voor de tijdsduur van de vlucht is de UTC aankomst- en vertrektijd bepalend.

Minder verbreid en bekend, maar net zo nuttig, is de verwijzing naar een *absolute locatie*. Wil men voorkomen dat in de verwijzingen naar locaties verschuivingen optreden (Luchthaven Praag ligt op het ene moment in Tsjecho-Slowakije, het volgende

moment ineens in Tsjechië) zou je kunnen denken aan een absolute plaatsaanduiding. GPS (Global Positioning System) coördinaten zijn absolute plaatsaanduidingen. GPS werkt met graden ooster- en westerlengte en zuider- en noorderbreedte en verfijningen daarvan. Routesystemen, kadastrale verwijzingen en verkeerssystemen gebruiken die coördinaten direct of indirect. Gezien de globalisering en het groeiende gebruik van satellieten moet een datamodelleerder notaties in absolute tijd en locatie tenminste overwegen.

Vroeger heette UTC ook wel GMT: Greenwich MeanTime. Maar Greenwich kent tegenwoordig ook zomer- en wintertijd, zodat deze aanduiding toch tot verwarring kon leiden.

Historie bewaren en corrigeren met terugwerkende kracht

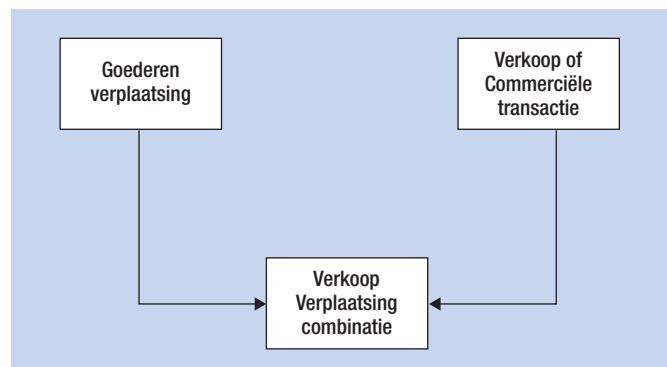
BI heeft als belangrijkste dimensie altijd de tijd. Inzicht in het wel en wee van de bedrijfsvoering doen managers in de allereerste plaats op de tijdsas. Hoe doet het bedrijf het nu ten opzichte van toen of straks? De tijdsas als aspect is zo evident dat ze als onderscheidend kenmerk van feiten vaak vergeten wordt. Bij BI is het tonen en bewaren van het juiste historisch perspectief rond een feit dan ook een niet te vermijden modelleringsvraagstuk. Er zijn veel manieren om inzichten in een tijdslijn te geven. En er zijn veel modelleringen te vinden die daarbij aansluiten. Er zijn ook *out-of-the-box* oplossingen om historie op te slaan. Of ze nu Kalido, BI-ready, Data Vault of anders heten, het zijn allemaal varianten van een methode die in tweedimensionale tabellen een derde dimensie maakt.

Een datamodelleerder moet notaties in absolute tijd en locatie tenminste overwegen.

Een verkoop wordt gedaan door een klant. Een klant heeft een adres. In de tijd gezien heeft die klant echter meer adressen, hij verhuist zo nu en dan. Er is dus een historische reeks klantversies. Een verkoopfeit hoort bij slechts één van die historische klantversies. Het kan zelfs zijn dat een wijziging plaatsvindt met terugwerkende kracht. In maart wordt bekend dat de werkelijke kosten van januari anders zijn. Dat is belangrijk voor de rapportage van januari. Die rapportage wordt dus met terugwerkende kracht inhoudelijk gewijzigd.

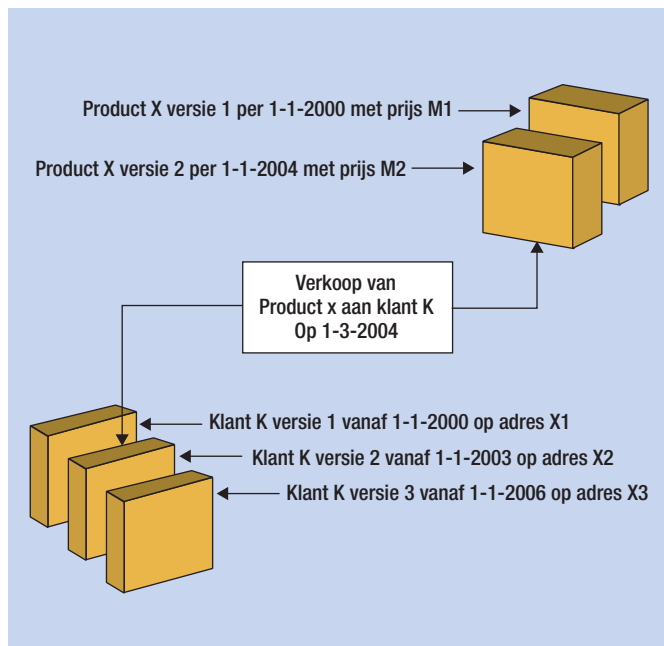
Veel nutsvoorzieningen werken op die manier. Er wordt een voorheffing gedaan. In de zomer wordt de eindafrekening bepaald aan de hand van de meterstanden. De echte energiekosten kunnen dan licht afwijken van de voorheffing.

Wanneer het mogelijk moet zijn rapporten te reproduceren volgens inmiddels verstreken inzichten (historie van historie), wordt die driedimensionaliteit zelfs vierdimensionaliteit. Als er



Afbeelding 1: Modelleringsvoorbeeld track & trace.

Thema Datawarehousing



Afbeelding 2: Bij een verkoop hoort een juiste historische versie van product en klant.

weinig tijd is voor een doordachte modellering en de bouw van een datawarehouse begint van *scratch*, is het aan te bevelen om naar een van die kant en klare producten te kijken. Een alternatief is de volgende oplossing.

Voorzie alle referentierecords waarvan je historie wilt bijhouden (bijvoorbeeld de verschillende adressen van een klant in de tijd) van een volgnummer. Wanneer je nu de juiste versie van de klant nodig hebt, kijk dan welk volgnummer een ingangsdatum heeft die zo groot mogelijk is, maar kleiner of gelijk aan de datum van het feit en je vindt de juiste referentie (zie afbeelding 2).

In SQL zou de query voor het zoeken naar de juiste klantversie met het juiste adres er zo uit kunnen zien:

```
Select klant.klantnr, klant.versienr, klant.adres
from klant,
      verkoop
where klant.klantnr = verkoop.klantnr
      and klantversienr = (select max(klant.versienr)
                          where ingangsdatum <= verkoop.verkoopdatum)
```

In geval van historie van historie moet je eerst in het referentieboek bladeren naar de juiste bladzijde met historieoverzichten. Dat is het equivalent van een view creëren die alle wijzigingen (invoerdatum) en volgnummers *na* de gewenste te reconstrueren rapportagedatum negeert. De gezochte historische lijnen negeren alle wijzigingen die na een bepaalde datum zijn ingevoerd, ook de wijzigingen met terugwerkende kracht. Die waren op de rapportagedatum immers nog niet bekend. Met die subset herhaal je de eerder beschreven stap en zoek je de historische waarde met de invoerdatum die kleiner of gelijk is aan de datum van het feit,

Newcom

From Data to Information to Knowledge

BI end-to-end oplossingen

- IT Management & IT Governance
- Woningbouwcorporaties
- Finance, HRM, Sales & Marketing
- Operational Performance Management

BI Projecten & Consultancy

- Resultaatgericht conform verwachting
- Productonafhankelijk, dus de beste keuze binnen budget
- Alles in één hand, van projectmanagement tot opleidingen en beheer
- Meer dan 20 consultants met ruime ervaring in BI & Data Warehousing

Interesse in onze dienstverlening?

Neem contact op met onze afdeling Sales & Marketing (sales@newcom.nl)

Optimaliseer uw informatievoorziening met Newcom Information Systems. Als expert op het gebied van Business Intelligence en Data Warehousing zorgen wij ervoor dat de informatievoorziening binnen een organisatie als een proces wordt gewaarborgd. Met hoogwaardige consultants voeren wij succesvolle projecten uit, zodat informatie binnen een organisatie bijdraagt tot winstgevendheid en effectiviteit.

Interesse om ons professionele team te versterken?

Neem contact op met Fiona de Jonge (fiona.de.jonge@newcom.nl)

Newcom Information Systems B.V.

P.O. Box 5631 T: +31 (76) 750 1800 info@newcom.nl
4801 EA Breda F: +31 (76) 750 1899 www.newcom.nl
Hoge Schouw 1G
The Netherlands



Klantnr	Klantversienr	Adres	Woonplaats	Ingangsdatum	Wijzigingsdatum
12	1	X1	Arnhem	1-1-2000	1-1-2000
12	2	X2	Utrecht	1-1-2004	1-1-2001
12	3	X3	Amsterdam	1-6-2001	1-1-2002
12	4	X2	Utrecht	1-1-2002	1-6-2002
12	5	X1	Arnhem	1-1-2003	1-1-2003
12	6	X4	Rotterdam	1-6-2003	1-6-2003
12	7	X5	Eindhoven	1-5-2003	1-1-2004
12	8	X6	Nijmegen	1-1-2004	1-6-2004
12	9	X7	Delft	1-1-2006	1-1-2006

maar zo groot mogelijk. De klantenrecords met terugwerkende kracht zouden er zoals in bovenstaande tabel uit kunnen zien.

Om de rapportage van 1 maart 2003 te reconstrueren moet eerst de historische lijn gereconstrueerd worden zoals die was op 1 maart 2003. Dat kan bijvoorbeeld met de volgende query:

```
Select * into werkklant
      from klant
Where wijzigingsdatum <= 1-3-2003
```

Daarmee ontstaat een werktabel met de records die verder onderzocht moeten worden. De volgnummers 6, 7, 8 en 9 staan daar niet in. Dat zijn wijzigingen van na de rapportagedatum. Record 2 staat er wel in, dat is immers op 1 januari 2001 al ingevoerd.

Vervolgens wordt de waarde van het transactiefit van bijvoorbeeld 1 maart 2002 weer getraceerd door het maximale volgnummer te selecteren, waarvoor geldt dat de ingangsdatum kleiner of gelijk aan 1 maart 2002. Klantversie 4 in dit geval; de klant woon-

de toen volgens de inzichten geldig op de rapportagedatum in Utrecht. De SQL kan ook in één samengestelde query met subselect worden vervat.

Het tonen en bewaren van het juiste historisch perspectief rond een feit is een niet te vermijden modelleringsvraagstuk

Nog een tip: bepaal en beschrijf duidelijk of de datum genoteerd wordt als *tot* of als *tot-en-met* datum. Het kan alle twee, als het maar duidelijk is en als de regel maar consequent gehanteerd wordt.

Karien Verhagen

Drs. C.W.J. Verhagen is senior BI-consultant/eigenaar van 4BIS Scholing en Advies.

Update

MicroStrategy en SPSS gaan samenwerken

MicroStrategy, leverancier van BI software, en SPSS, leverancier van software-oplossingen voor predictive analytics, hebben een strategische samenwerkings-overeenkomst gesloten. De overeenkomst brengt de BI-technologie van MicroStrategy samen met predictive analytics van SPSS, en omvat business development en het opzetten van gezamenlijke marketingprogramma's. Met behulp van PMML (predictive modeling mark-up language) integratie

kunnen gebruikers voorspellende analyse-algoritmes beïnvloeden door het gebruik van MicroStrategy-software. De technologie van SPSS is goed schaalbaar en samen te gebruiken met de producten van MicroStrategy. Meer informatie op www.spss.nl

Nieuwe Database Designer voor PostgreSQL

MicroOLAP bracht onlangs Database Designer voor PostgreSQL versie 1.0.6 op de markt, een Windows applicatie voor het bouwen en onderhouden van deze

databases. De nieuwe versie ondersteunt alle PostgreSQL syntax-extensies en alle typen tabellen, referenties, triggers, stored procedures, functies en velden. Bovendien kunnen bestaande databases, zoals MS Access, Sybase, Oracle, Informix, MSSQL en DBF worden geïmporteerd.

Database Designer for PostgreSQL heeft een directe verbinding met PostgreSQL, zonder dat ODBC of ADO/OLEDB als intermediair nodig zijn. Het ondersteunt PostgreSQL 7.x en nieuwer. Meer informatie op <http://microolap.com>