

Met tools, ontwikkelmethode en architectuur  
behoren lange ontwikkeltijden tot het verleden

# ‘Kant-en-klare’ datamarts



Freek Kamst

**Stel je eens voor: een datawarehouseproject waarin het beoogde profijt al zichtbaar is binnen een maand. Dat is vast te mooi om waar te zijn. Of toch niet...? Freek Kamst**

**duidt de trend richting kant-en-klare datamarts per bedrijfstak. Die komen het beste tot hun recht in een ‘prisma-architectuur’ en met een iteratieve ontwikkelmethode.**

Steeds meer organisaties zien de potentiële voordelen van een datawarehouse of datamart. Daardoor neemt het aantal projecten aanzienlijk toe. Tot nog toe kon slechts een selectief aantal ondernemingen zich dergelijke projecten veroorloven. Maar het aanbod van standaard applica-

ties voor datamodellering en het ontwerpen van datawarehouses en datamarts groeit. Dat brengt nu ook voor middelgrote bedrijven de mogelijkheden van business intelligence binnen handbereik. Een nog jongere trend sluit zich hierbij aan: er komen standaard datamodellen voor specifieke processen binnen diverse bedrijfstakken op de markt. Hierdoor ontstaan min of meer kant-en-klare datamarts die direct gevuld kunnen worden met de gegevens uit de bronssystemen. De marketingafdelingen van deze leveranciers scheiden slogans af als *datamart in a box* en *instant datawarehouse*, waardoor de suggestie wordt gewekt dat het bouwen van een datawarehouse of datamart snel en eenvoudig te realiseren is. Toch moeten deze tools nog steeds als gereedschappen worden gezien en vraagt het bouwen van een datawarehouse-omgeving -een onomkeer-

baar proces- de nodige organisatorische inzichten en automatiseringskennis. Wel kan het daardoor een minder langdurige en kostbare zaak worden.

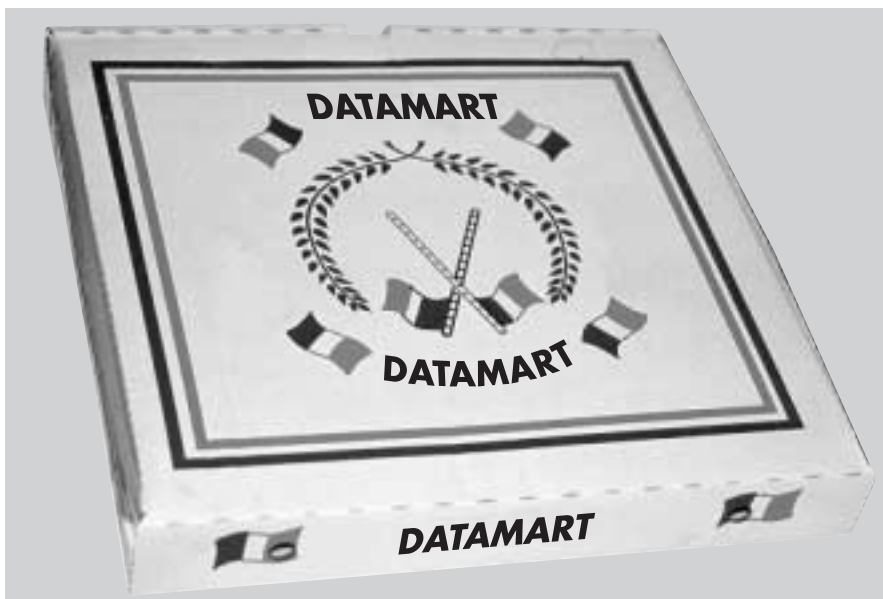
## WAREHOUSE OF MART?

De meeste organisaties realiseren zich de voordelen van één datawarehouse waarin alle bedrijfsgegevens aanwezig zijn, het zogenaamde enterprise datawarehouse. Onder druk van de korte-termijnbehoefte aan informatie wordt meestal echter al snel gekozen voor subjectgeoriënteerde datamarts. Deze vullen inderdaad op korte termijn de gewenste informatiebehoefte in, maar zijn op lange termijn nadelig, omdat hun doorgaans *onafhankelijke* opzet ze moeilijk integreerbaar maakt.

## Top-down

Er is al veel gepubliceerd over de verschillen tussen datamarts en datawarehouses. Specialisten stemmen over het algemeen in met de opvatting dat een enterprise datawarehouse-omgeving bestaat uit één datawarehouse met een of meer *afhankelijke*, subjectgeoriënteerde datamarts. Deze top-down-benadering heeft de volgende voordelen.

- De datawarehouse-omgeving wordt eenduidig opgezet, met standaard methoden en procedures voor het ontwerpen en ontwikkelen van data-extracties, transformaties en eindgebruikerapplicaties.
- Er ontstaat een datawarehouse-omgeving waarin men op eenvoudige wijze



een organisatiebreed overzicht kan krijgen van de bedrijfsprocessen en hun prestaties. Dat geeft meer mogelijkheden om de ontwikkeling van de bedrijfsvoering optimaal te benutten.

- Het datawarehouse is opgezet naar een organisatiebreed, subjectgeoriënteerd informatiemodel. Problemen omtrent gegevensintegratie worden daardoor geminimaliseerd.
- Deze benadering opent de optie om de onderliggende datamarts te construeren vanuit het gegevensmodel van het datawarehouse.

Maar er kleven natuurlijk ook nadelen aan deze top-down-benadering.

- De ervaring leert dat IT-gedreven top-down-projecten vaak leiden tot zeer lange doorlooptijden, hoge kosten en een zwakke eindgebruikersfunctionaliteit in het begin; zelfs wanneer een uitgebreide voorcalculatie op het project is uitgevoerd.
- De gewenste informatie laat te lang op zich wachten bij de ontwikkeling van een enterprise datawarehouse. De informatiebehoefte zal tijdens de ontwikkeling meerdere malen veranderen, waardoor de ontwikkelcyclus altijd achter de feiten aan zal lopen.

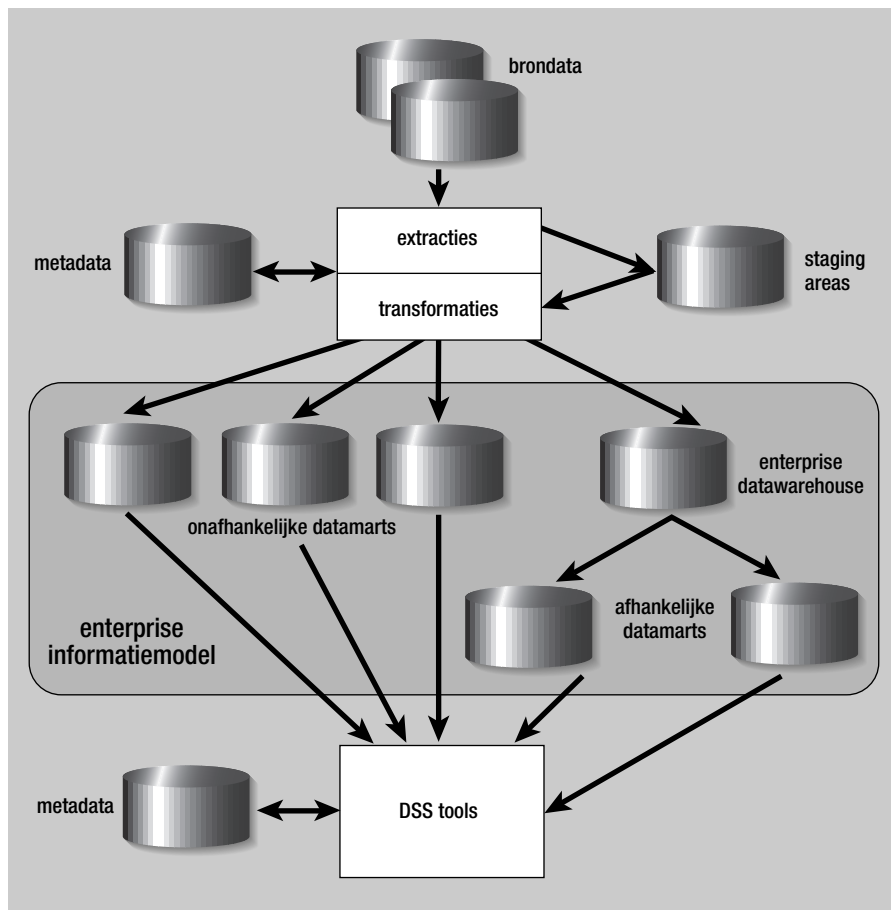
**Bottom-up**

Sinds een aantal jaren doet een alternatieve oplossing opgeld, om deze nadelen te ondervangen. Daarbij ontwikkelt men eerst de datamarts en construeert vervol-

***Bouwen van een datawarehouse kan wel korter en goedkoper, maar niet zonder inzicht en kennis - tools zijn en blijven gereedschappen***

gens vandaaruit het datawarehouse.

Sommige voorstanders van deze bottom-up-benadering beargumenteren dat een enterprise datawarehouse niet eens noodzakelijk is. Een ondernemingswijde blik zou je namelijk kunnen verkrijgen door het opzetten van een gedistribueerde data-



FIGUUR 1: EEN PRISMATISCHE DATAWAREHOUSE-OMGEVING.

warehouse-omgeving, met behulp van databasemiddleware. Maar ook daaraan zit weer een aantal nadelen vast.

- Door een ongecoördineerde ontwikkeling van allerlei datamarts ontstaan problemen met gegevensintegratie. Sterk aanwezig is bovendien het risico dat visie ontbreekt om door te groeien naar een uiteindelijke situatie. Deze problemen zijn overigens te voorkomen door het vooraf definiëren van een enterprise-informatiemodel.
- Wanneer de datamarts gedurende een periode operationeel zijn, ontstaan al snel behoeften bij andere afdelingen om bijvoorbeeld cross-business-analyses te maken en dus meerdere datamarts te raadplegen. Gelijkijdig benaderen van verschillende datamarts is echter uitermate complex zonder de juiste databasemiddleware. Nadeel is dat deze middleware moeilijk is te beheren en meestal leidt tot slechte systeemprestaties.
- Vaak zetten de aanwezige gegevens in

de operationele systemen aan tot snel en frequent ontwikkelen van datamarts. Toch zou het ontwerp moeten worden opgesteld aan de hand van de informatiebehoeften van de eindgebruikers.

**PRISMATISCH DATAWAREHOUSE**

Het is duidelijk dat zowel 'top-down' als 'bottom-up' evenveel voor- als nadelen met zich omdragen. De meeste organisaties willen natuurlijk een datawarehouse-toepassing met de laagste kosten en de kortste terugverdientijd. Om dit te bereiken moet het ontwerp van de eventuele datamarts gebaseerd zijn op een enterprise-informatiemodel. Met andere woorden: organisaties moeten het ontwerp van hun datamarts opstellen aan de hand van een enterprise datawarehouse-datamodel. Deze hybride oplossing bevat het beste van beide genoemde benaderingen en kan worden ondersteund door de gangbare ite-

# Globale ontwikkeltijden bij een middelgrote datamart

Onderstaande tabel geeft een situatie weer van de globale ontwikkeltijden bij een middelgrote datamart. Het aantal dagen in de eerste kolom is gebaseerd op een ontwikkeling vanaf scratch zonder gebruik te maken van tools. De tweede kolom geeft het aantal dagen weer van dezelfde situatie, maar met de inzet van ontwerp- en extractie-tools en een standaard datamodel.

Activiteit	zonder tools	met tools	Opmerking
Readiness assessment (haalbaarheidsonderzoek)	30	15	
Requirements gathering (informatie-inventarisatie en -analyse)	120	16	(1)
Opstellen logisch datamodel	240	0	(2)
Opstellen fysiek database-ontwerp	360	8	(1)
Ontwikkeling DDL voor creëren tabellen en indexen	3	1	(1)
Opstellen mappings van bron naar datawarehouse	480	120	(3)
Ontwerpen eindgebruikersapplicaties	160	33	(4)
Ontwerpen extracties (ETL)	60	36	(3)
Testgegevens genereren	80	3	(5)
Testen laadprocedures	10	8	(3)
Ontwikkeling eindgebruikersapplicaties	360	40	(4)
Ontwikkeling extractieprogrammatuur	240	56	(3)
Testen extractieprogrammatuur	60	20	(3)
Testen applicaties met testdata	160	4	(3/5)
Initieel laden van de data	80	30	
Testen met productiedata en tuning datawarehouse	240	40	
Projectmanagement	240	60	
Datawarehouse-architect	90	20	
Totaal	3013	510	

## Opmerkingen

1. Het standaard informatie- en datamodel is gebaseerd op diverse succesvolle projecten in specifieke bedrijfsprocessen. Aanpassen van deze modellen naar de specifieke wensen en eisen van de organisatie bespaart veel tijd in het analyse- en ontwerptraject.
2. Ontwerp van het logisch datamodel vraagt nauwelijks enige inspanning, omdat dit direct wordt aangeleverd.
3. Gebruik van ETL-tools bespaart ongeveer een derde van de benodigde inspanningen.
4. Een aantal leveranciers verkoopt standaard eindgebruikersapplicaties bij het specifieke proces; tevens worden de metadata voor diverse eindgebruikersapplicaties automatisch gegenereerd.
5. Testgegevens zijn automatisch te genereren.

ratieve ontwikkelmethoden. Deze architectuur bezit de vorm en de aspecten van een prisma. Daarom duiden we deze architectuur hier verder aan als *prismatische datawarehouse-omgeving*. De sleutel tot gegevensintegratie in het prismatische datawarehouse is altijd het enterprise-informatiemodel. Dat is uniek binnen de organisatie en vormt tevens de bron van

alle metadata. Prismatisch datawarehouse en enterprise-informatiemodel maken de integratie van de diverse datamarts ook veel eenvoudiger. Bij het bouwen van nieuwe datamarts is de ontwikkeling van de data-extracties en de bijbehorende transformaties een vast terugkerend proces. In traditionele datawarehouse-omgevingen nemen de aantallen extracties en

transformaties toe naar gelang er datamarts bijkomen. Inherent daaraan wordt het onderhoud op deze data-extracties steeds intensiever, terwijl de coördinatie van het laadproces almaar ingewikkelder wordt.

In een prismatische omgeving koppelen data-extracties los van de transformaties en ontwikkelt men een beperkt aantal gemeenschappelijke routines die de brongegevens in een gemeenschappelijke *staging area* laden. Vandaaruit worden de gegevens aan de diverse datamarts en/of het datawarehouse geleverd. Op dat moment vinden transformaties plaats. Moeten nieuwe datamarts of datawarehouse-deelgebieden worden toegevoegd, dan kunnen eenvoudig de bestaande extractie-applicaties en transformatieroutines worden gebruikt of aangepast. Deze methode werkt uitstekend in een prismatische datawarehouse-omgeving, waarbij het enterprise-informatiemodel als basis dient voor het ontwerp van de extractie-applicaties en de inrichting van de *staging area*.

## ONTWIKKELING EN IMPLEMENTATIE

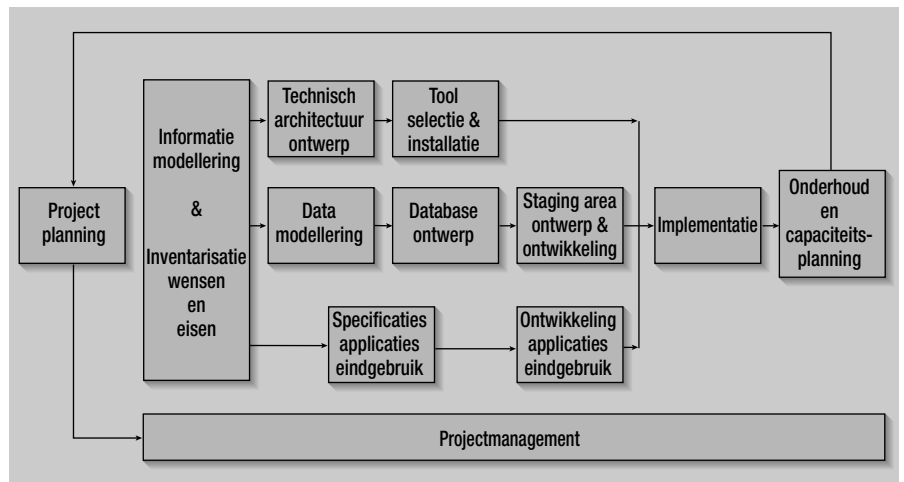
Hoewel een prismatische datawarehouse-omgeving een organisatiebrede oplossing is, hoeft deze niet per se in één keer te worden geïmplementeerd. Projecten die dergelijke *big bang-operaties* voor ogen hadden, zijn nagenoeg alle mislukt. Een iteratieve ontwikkelmethode, waarbij vaste deelgebieden (incrementen) worden ontwikkeld en geïmplementeerd, hebben tot nu toe de beste en snelste resultaten geleverd. Op veel succes in de praktijk kan *The Business Dimensional Life Cycle* van Ralph Kimball bogen. Deze methode voor datawarehouse-projectmanagement zorgt ervoor dat je de diverse activiteiten en componenten van een warehouse-increment op de juiste manier uitvoert en ten slotte tijdig integreert. Benadrukt moet worden dat niet alleen het perfecte informatiemodel of de aanschaf van geavanceerde tools tot succes leidt. Op de juiste manier dirigeren van een warehouse-projectteam is een van de

belangrijkste aspecten. Toch valt de eerste datamart of het eerste datawarehouse-increment niet te onderschatten. Ook wie een iteratieve ontwikkelmethode volgt moet zich nog behoorlijk wat inspanning getroosten voordat de eindgebruikers de eerste resultaten zien. Bij het eerste deelgebied worden immers de meeste dimensietabellen gedefinieerd en moet de staging area worden gebouwd.

Uit ervaringen die inmiddels met de diverse datawarehouseprojecten zijn opgedaan, is bekend dat tachtig procent van de informatie- en datamodellen binnen specifieke processen van bepaalde bedrijfstakken sterk op elkaar lijken. Een aantal leveranciers van ontwerp- en extractie-tools heeft deze constatering aangegrepen om bij hun producten standaard datamodellen voor deze specifieke processen aan te bieden. Het resultaat is verbluffend! Door een vijfde deel op een standaard datamodel aan te passen aan de wensen en eisen van de organisatie wordt de ontwikkeltijd van een datamart of datawarehouse-subject enorm verkort. In het kader *Globale ontwikkeltijden bij een middelgrote datamart* is een situatie weergegeven.

**LEVERANCIERS**

Een aantal leveranciers van ontwerp- en ETL-tools integreert inmiddels standaard datamodellen voor specifieke verticale bedrijfsprocessen in hun producten of levert ze als extra optie. We geven een korte beschrijving van een *low-end* (AppsMart) en een *high-end* (Industry



FIGUUR 2: 'BUSINESS DIMENSIONAL LIFE CYCLE'.

Warehouse Studio) -toepassing waarin het gebruik van verticale standaard informatiemodellen is geïntegreerd.

**AppsMart**

AppsMart is een rapid application development (RAD) -tool voor het ontwerpen en bouwen van datamarts die goed in een prismatische datawarehouse-omgeving implementeerbaar zijn. Een van de voordelen van AppsMart, een product van de Britse AppsCo, is dat met de inzet van standaard datamodellen (templates) snel een enterprise-informatiemodel te ontwerpen is. Uitgangspunt voor de verdere ontwikkeling van een datamart/datawarehouse-omgeving is dit informatiemodel. AppsMart bestaat uit drie hoofdmodules: Scratchpad Editor, Template Designer en AppsMart Repository.

Scratchpad Editor zorgt voor de ontwikkeling van het informatiemodel, waarin onder meer wensen en eisen van de eindgebruikers worden vastgelegd. AppsMart baseert het ontwerp van datamarts op hun gewenste informatiebehoefte, en niet op de aanwezige brongegevens in de transactionele systemen. Een dergelijke benadering voldoet beter en sneller aan het gewenste verwachtingspatroon.

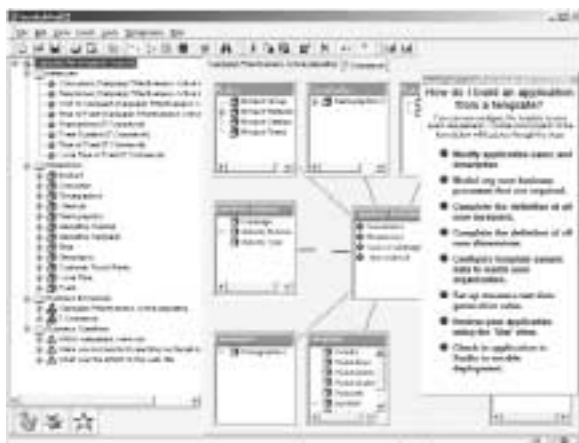
Het informatiemodel wordt opgeslagen in een template. De ontwikke-

laar/analist kan deze zelf ontwerpen of een standaard kiezen en die aanpassen. Via wizards en een gebruikersvriendelijke point-and-click-interface is het mogelijk elementen aan te passen, zodat het gewenste informatiemodel snel tot stand komt. De businessprocessen binnen het informatiemodel worden conform het dimensionele model opgesteld. Dimensionele modellen zijn standaard in een datawarehouse-omgeving.

*Projecten die 'big bang'-operaties voor ogen hadden, zijn nagenoeg alle mislukt*

Met de Template Designer kunnen partners van AppsCo specifieke templates ontwikkelen in marktsegmenten waarin zij ervaring hebben. Datawarehouse-ontwikkelaars gebruiken deze designer voor beheer en onderhoud van het enterprise-informatiemodel van bijvoorbeeld een prismatische datawarehouse-omgeving. Tijdens de ontwikkeling van een nieuwe datamart kunnen de gegevens worden geïntegreerd in dit model. De aangepaste template kan vervolgens weer worden gebruikt door AppsMart Studio.

AppsMart Studio acteert als repository voor het beheren van de templates. Door versiebeheer blijft de integriteit daarvan gewaarborgd. AppsMart Studio zorgt vervolgens ook voor het genereren van de databaseschemadefinities en de metadata



FIGUUR 3: STANDAARD DATAMODEL IN APPSMART.

voor het business intelligence-tool.

Ook zijn automatisch testdata te genereren, die vervolgens in de datamart kunnen worden geladen.

AppsMart is gebaseerd op integratie met Microsoft-producten. Dat maakt business intelligence voor middelgrote organi-

### Tachtig procent van de informatie- en datamodelen binnen specifieke processen van bepaalde bedrijfstakken is gelijk

saties mogelijk. Maar het product wordt ook ingezet bij het definiëren van enterprise-informatie- en datamodelen bij omvangrijke datawarehouseprojecten.

#### Industry Warehouse Studio

Industry Warehouse Studio van Sybase bestaat uit een kant-en-klare infrastructuur voor datawarehousing. Componenten zijn een verticaal georiënteerd informatiemodel, een databasestructuur, een methode, applicaties, metadata, een tool om de metadata te beheren en een om het informatiemodel aan te passen aan de specifieke organisatie-omgeving. Industry Warehouse Studio is tevens integreerbaar met bestaande omgevingen, zodat verder gebouwd kan worden op eenmaal gedane technologische investeringen.

Banken, verzekeringen, telecommunicatie en gezondheidszorg zijn onder andere de verticale marktsegmenten die momenteel worden voorzien van standaard datamodellen. Maar zo is er ook een standaard datamodel voor 'customer relationship marketing'. Dit Sybase-product vereist geen Sybase-database. Dat biedt de organisatie de mogelijkheid verder te bouwen op bestaande omgevingen, waarin bijvoorbeeld al een database, business intelligence of ETL-tool aanwezig is.

Industry Warehouse Studio bevat twee modules voor beheer en ontwerp van een datawarehouse-omgeving: Sybase Warehouse Architect en Sybase Warehouse Control Center. De eerste is een tool voor datawarehouse-ontwerp op conceptueel en fysiek niveau. Het biedt de mogelijkheid

een multidimensionele omgeving te ontwerpen, beheren en documenteren. Ook het genereren van frontend-metadata voor onder andere Cognos, MicroStrategy, Oracle Express, Sybase PowerDimensions en Hyperion Essbase is een functionaliteit van deze module. Databaseschema's van 65 verschillende databases en versies kunnen worden ingelezen. Wanneer een datawarehouse-ontwerp gereed is, kan de module de DDL-scripts genereren waarmee de database -naar eigen keuze- is te creëren.

De module Warehouse Control Center is een metadata-beheertool. Hiermee kunnen de metadata van een datawarehouse -technische, business-, contextuele en inhoudelijke metadata- beschikbaar worden gemaakt aan ontwikkelaars en eindgebruikers. De module importeert de technische en business-metadata direct vanuit Warehouse Architect. Hierdoor is het niet nodig een metadata-structuur op te zetten en verloopt de ontwikkeling van bijvoorbeeld een prismatische datawarehouse-omgeving sneller. Bij het laden van de metadata vindt een controle plaats, zodat eindgebruikers zeker kunnen weten dat ze de juiste gegevens aanschouwen.

### Uiteindelijk transformeert ook de datawarehousemarkt naar goedkopere standaard oplossingen

Warehouse Control Center is ook in staat metadata-omgevingen te creëren voor diverse frontend-tools, zoals Business Objects en Cognos Impromptu. Door deze functionaliteit kan de eindgebruiker bijvoorbeeld gewenste entiteiten en attributen selecteren en vervolgens automatisch een *universe* (BO) of *catalogue* (Impromptu) genereren.

Naast de tools zijn er templates voor de standaard datamodellen. De datastructuur is modulair opgezet, met als uitgangspunt een kerndatamodel van ongeveer zeventig tabellen. Deze tabellen zijn identiek voor elk verticaal model. Per 'verticale' bedrijfstak is er vervolgens een op het kernmodel

aanluitend datamodel. Dat datamodel bestaat uit modulair opgezette aggregaat-, feedback-, externe en toolspecifieke tabellen. Bij Industry Warehouse Studio leverbare verticale sectoren zijn:

- CRM (campaign management analysis, customer profiling, customer care analyses, customer loyalty, sales analyses);
- Retail Banking;
- Credit Card Business
- Verzekeringen (levensverzekeringen, goederen, ongevallen);
- Telecommunicatie;
- Gezondheidszorg.

Sybase levert hiermee niet alleen de tools, maar ook het concept om een datawarehouse-omgeving te bouwen en te beheren.

#### PERSPECTIEF

AppsCo en Sybase hebben de signalen uit datawarehouse- en datamartprojecten goed begrepen en dat biedt perspectief. Zij komen conceptueel met toepassingen die de traditionele risico's van dergelijke projecten verkleinen of zelfs elimineren. Wie ze goed gebruikt, met een iteratieve ontwikkelmethode en in een prismatische datawarehouse-architectuur, zal er profijt van trekken. De lange ontwikkeltijden van datawarehouses zullen weldra tot het verleden behoren. De verwachting is dan ook dat de overige leveranciers van tools voor business intelligence en datawarehousing gaan aanhaken bij deze trend. Voor specifieke bedrijfstakken en markten zullen veelal de partners van de leveranciers met kant-en-klare oplossingen komen. Daardoor transformeert uiteindelijk ook de datawarehousemarkt naar goedkopere standaard oplossingen. ●

#### Referenties:

Colin J. White: *Managing Data Mart Development*. Database Associates International Inc. 1998.  
R. Kimball, L. Reeves, M. Ross, W. Thornthwaite: *The Data Warehouse Lifecycle Toolkit*. J. Wiley & Sons, 1998.

Freek Kamst (freek.kamst@newcom.nl) is senior consultant bij Newcom Information Systems.