

Laad- en transformatiefaciliteit maakt direct PowerPlay-kubussen aan

# Mercedes vult datawarehouse met ETL-tool van Cognos

Robbert Hoeffnagel

**B**ij het Mercedes Benz Customer Assistance Center (CAC) in Maastricht worden alle contacten met Europese gebruikers van Mercedes-voertuigen vastgelegd. Dat kan variëren van een melding van een pechgeval tot een aanvraag voor een brochure voor een nieuw model. Om de gegevens over inmiddels ruim 1,7 miljoen klanten te kunnen ontsluiten, werden tot voor kort analyses rechtstreeks op de operationele systemen uitgevoerd. Inmiddels werkt men hard aan de opbouw van een datawarehouse, waarbij DecisionStream als ETL-tool en PowerPlay-kubussen voor analysedoeleinden een belangrijke rol spelen.

Dagelijks komen honderden gesprekken, faxen en e-mails binnen bij het Mercedes Benz Customer Assistance Center (CAC) in Maastricht. Het gaat om vragen en meldingen over tal van onderwerpen. Van een melding naar aanleiding van een ongeval en een aanvraag van kinderen die sprekebeurten over auto's of dit specifieke auto-merk willen houden tot vragen over nieuw geannonceerde modellen.

Voor het prestigieuze Duitse automerk zijn deze contacten van groot belang. Serviceverlening is een belangrijk hulpmiddel om bestaande klanten te behouden en nieuwe te winnen. Huidige rijders zijn in veel gevallen afhankelijk van de ondersteuning die het CAC bij een technisch probleem kan bieden. Bovendien handelt het centrum veel contacten af met potentiële nieuwe afnemers. Daarom worden

alle contacten gestructureerd vastgelegd. Daarvoor gebruikt het CAC de callcentre-programmatuur van Siebel. De cases worden vastgelegd in de onderliggende Oracle-database. Aangezien het CAC voor elf Europese landen actief is, gaat het om grote aantallen cases en verslagleggingen. In de database zijn inmiddels gegevens over ruim 1,7 miljoen klanten en auto's verzameld. Het aantal geregistreerde telefoongesprekken ligt op 2 miljoen per jaar.

## EUROPEES SERVICECENTRUM

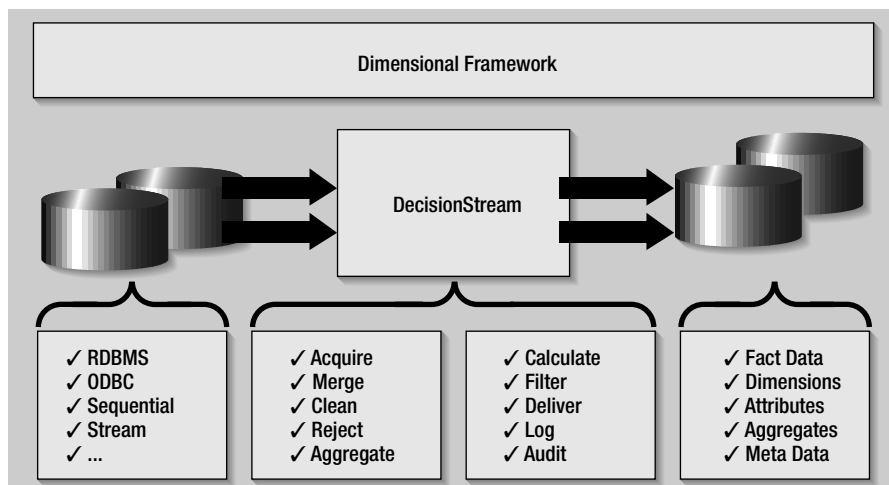
Het CAC is ontstaan door een geleidelijke samenvoeging van de callcentres die de diverse landenorganisaties van Mercedes in Europa in gebruik hadden. Deze zijn één voor één overgebracht naar Maastricht, vertelt Ingrid Posthuma, destijds als research & planning manager verantwoordelijk voor deze operatie. "Wij willen iede-

re klant, ongeacht waar een melding vandaan komt, op dezelfde manier kunnen aanspreken en verder helpen. Bovendien

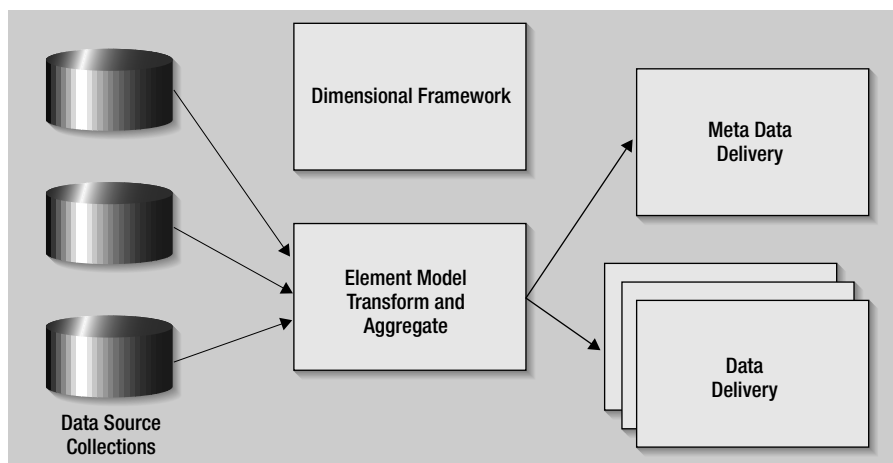
*"Onze vorige aanpak bood onvoldoende flexibiliteit en grafische ondersteuning"*

willen we dat de serviceverlening altijd en overal van hetzelfde hoge niveau is.

Daartoe kunnen alle Europese klanten in geval van problemen nu één centraal telefoonnummer bellen dat hen in contact brengt met het callcentre, hier in Limburg. Wij zien waar het gesprek vandaan komt en veelal ook direct in welke taal de persoon in kwestie aangesproken dient te worden. De callcentremedewerker die het gesprek aanneemt, schakelt vervolgens die dealer in die zich fysiek het dichtst bij de



FIGUUR 1: OVERZICHT VAN DE ARCHITECTUUR VAN DECISIONSTREAM VAN COGNOS.



FIGUUR 2: RELATIE TUSSEN DECISIONSTREAM DESIGNER EN DECISIONSTREAM ENGINE.

klant bevindt of het beste voor ondersteuning is toegerust. Deze dealer meldt na afhandeling van de melding vervolgens weer aan het CAC wat precies het probleem was en voor welke oplossing is gekozen."

De diverse verzoeken om assistentie en informatie worden via Siebel Call Center in Oracle vastgelegd en verwerkt. Jaarlijks gaat het om circa een half miljoen cases. Gemiddeld telt ieder case vijf contacten met de klant. Een dergelijke database bevat veel interessante gegevens, zowel voor de centrale Mercedes Benz-organisatie als voor de diverse landenorganisaties. Enerzijds is het gewenst om ter referentie alle cases te bewaren, zodat in voorkomende gevallen teruggezocht kan worden wat er precies is gebeurd. Daarnaast bieden de gegevens belangrijke informatie over de mate waarin en de wijze waarop dealers met meldingen omgaan.

## CENTRAAL DATAWAREHOUSE

De vraag is alleen hoe dit soort informatie op een efficiënte manier aan de database-omgeving te onttrekken valt. Het creëren van een datawarehouse lijkt de meest voor de handliggende aanpak. Dat is dan ook de weg die het CAC inmiddels volgt. Het centrum is een zelfstandige organisatie binnen Mercedes Benz. Posthuma: "Dat betekent dat we een krachtige analyse- en rapportagevoorziening nodig hebben voor zowel de dienstverlening aan onze

opdrachtgevers -de landenorganisaties- als voor het kunnen beoordelen en verbeteren van onze interne manier van werken."

In eerste instantie werden analyses en rapporten direct op basis van de gegevens in de operationele systemen gemaakt. "Dat gebeurde onder andere door data apart te zetten in Microsoft Access en vervolgens met Crystal Reports analyses en rapporten te maken", vertelt Mark Gorissen, als specialist in research and planning werkzaam bij CAC's afdeling Facility Services. "Dat was uiteindelijk echter geen goede aanpak. Het bood geen flexibiliteit en het ontbrak bijvoorbeeld aan voldoende mogelijkheden om grafische presentaties van de

*"De tools van Oracle zijn op Java gebaseerd, en daarover hebben wij te weinig kennis in huis"*

analysegegevens te kunnen vervaardigen. Het was bovendien de bedoeling dat de diverse landenorganisaties zelf hun analyses konden maken, maar dat lukte op die manier niet goed. Met als gevolg dat wij hier zelf de analyses maakten en de daarop gebaseerde rapporten vervaardigden. Daarmee begonnen wij steeds meer de rol op ons te nemen van een soort printshop voor de diverse landen."

Daarom werd besloten een centraal datawarehouse aan te leggen, compleet met alle daarbij behorende hulpmiddelen voor analyse en rapportage, en waarin

gegevens worden verzameld uit tal van systemen, waaronder het callcentre, maar ook financiële toepassingen als Exact en andere operationele systemen. Vervolgens is onder meer onderzoek gedaan naar de mogelijkheden die het bestaande duo Siebel en Oracle op dit terrein te bieden heeft.

Posthuma: "Dat bleek naar onze mening toch niet echt te voldoen. De aanwezige mogelijkheden voor het maken van rapporten bleken te beperkt. Daarbij ontbrak het bovendien aan de flexibiliteit die wij zochten. Bovendien bestaat binnen de onze organisatie een duidelijke voorkeur voor bepaalde producten als het om analyse en rapportage gaat, en dat zijn Cognos-producten. Daarbij kwam dat de hulpmiddelen die Oracle biedt voor het verzamelen van gegevens uit andere systemen op Java gebaseerd zijn. Hierbij wordt veel met zelf te schrijven scripts gewerkt en daarover hebben wij weinig kennis in huis."

## RAPPORTAGE-BEHOEFTEN

Uiteindelijk is gekozen voor een aanpak waarbij de zestig betrokken managers en coördinatoren via een webinterface zelf rapportages kunnen maken op basis van de in Oracle vastgelegde gegevens. Hiervoor wordt Cognos PowerPlay Web gebruikt. Uit intern onderzoek werd duidelijk aan welke rapportages behoefte was en konden in PowerPlay kubussen worden opgezet. Om gebruikers te laten wennen aan het zelf kunnen maken van analyses, is een getrapte aanpak gevolgd, waarbij langzaam maar zeker de vrijheid van handelen werd vergroot. Zo is onder meer deze tussenstap gemaakt: kleine reeksen van standaardrapportages werden in de vorm van PDF-bestanden aangeboden.

Het vullen van het onderliggende en op een Oracle-database gebaseerd datawarehouse gebeurt eveneens met een Cognos-product: DecisionStream.

## ETL-TOOL

DecisionStream is een ETL-hulpmiddel voor extractie, transformatie en het vervol-

gens laden van grote hoeveelheden data uit meerdere systemen in multidimensionale databases. Het tool maakt het mogelijk dat een gegevensstroom op gang komt tussen het bronsysteem en het doelsysteem. Aangezien zelden data rechtstreeks in een doelsysteem kunnen worden ingelezen, is een bewerkingsslag nodig. De hiervoor benodigde transformatieregels zijn met DecisionStream op te stellen en vast te leggen in een repository met metadata.

### 'DIMENSIONAL FRAMEWORK'

De repository met metadata -in het jargon van Cognos 'catalog' geheten- is in iedere relationele database onder te brengen. Bovendien kan deze catalog worden gedeeld door meerdere engines, zodat meerdere datawarehouses hiervan gebruik kunnen maken. De in de catalog vastgelegde transformatieregels zijn bedoeld voor het beschrijven van een *dimensional framework*. Dit raamwerk beschrijft de dimensies die door de transformatie-engine kunnen worden gebruikt, zoals klant, product en plaats. Het ondersteunt hierbij een reeks van standaard hiërarchieën, waaronder asymmetrische. Ook is het gebruik van 'shared members' mogelijk en wordt het concept van individuele attributen van dimensies op verschillende niveaus in de hiërarchie ondersteund.

Ieder DecisionStream 'build object' of 'build' bestaat uit een aantal processen en stappen bedoeld voor het verzamelen van gegevens, het transformeren van deze data, de daaropvolgende aggregatieslag en het uiteindelijk leveren van de gegevens aan het doelsysteem. Aangezien dit proces van verzamelen en bewerken van data veelal in beperkte tijdvensters moet plaatsvinden, worden technieken als smart caching en memory management toegepast om de 'throughput' zoveel mogelijk te verhogen. Doel van dit soort technieken is de bewerkingen zoveel mogelijk in het geheugen van databaseserver te laten plaatsvinden, zodat het meermalen laden en weer wegschrijven van gegevens zoveel mogelijk wordt voorkomen. Het effect hier-

van hangt uiteraard sterk af van de specifieke omstandigheden.

Het bronsysteem wordt geselecteerd aan de hand van een 'data source definition'. Hierin kunnen tevens de regels voor eventuele cleansing-operaties worden vastgelegd, zodat de data zijn samen te voegen met de in het dimensional framework opgenomen gegevens. In principe worden data ingevoerd op het laagste niveau van de hiërarchie, maar het is eventueel ook mogelijk gegevens te accepteren op hogere niveaus.

De programmatuur ondersteunt een reeks van operaties voor aggregatie en transformatie. Denk aan string-verwerkingen, numerieke functies, manipulatie van datumvelden, conditionele bewerkingen,

*"Wij moeten iedere dag kunnen beschikken over volledig bijgewerkte gegevens, en dus iedere nacht updaten"*

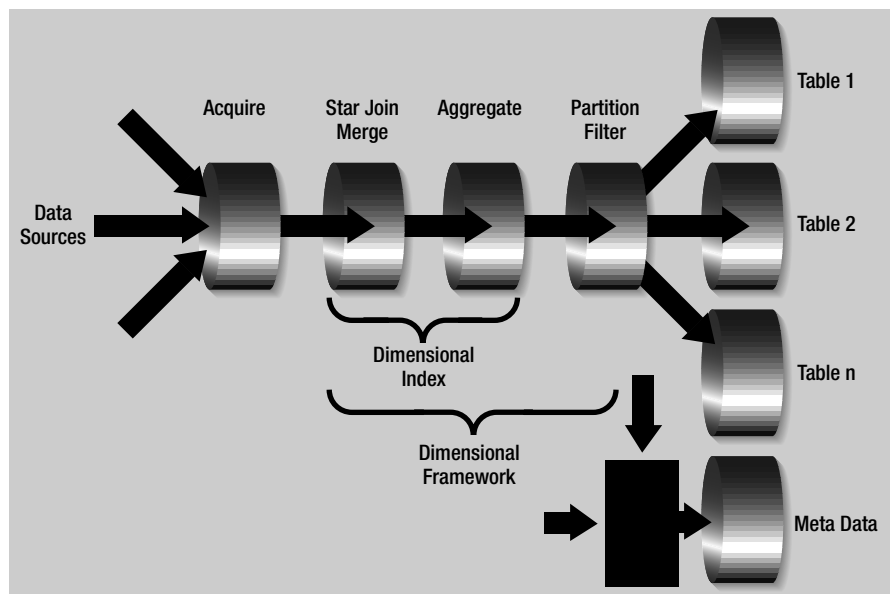
functies die afhankelijk zijn van de hiërarchie, het automatisch genereren van *surrogate keys* en bewerkingen die gerelateerd zijn aan attributen. Hierbij worden tevens standaard gegevenstructuren als Snowflake- en Star-schema's ondersteund.

Bij het leveren van de resulterende gegevens aan het datawarehouse (het doelsysteem) is de data stream te splitsen in partities, waardoor een parallelle en snellere aanlevering mogelijk wordt. Partities kunnen door middel van filters worden gerealiseerd en beheerd. Ook zijn gedistribueerde aggregatiemodellen te hanteren.

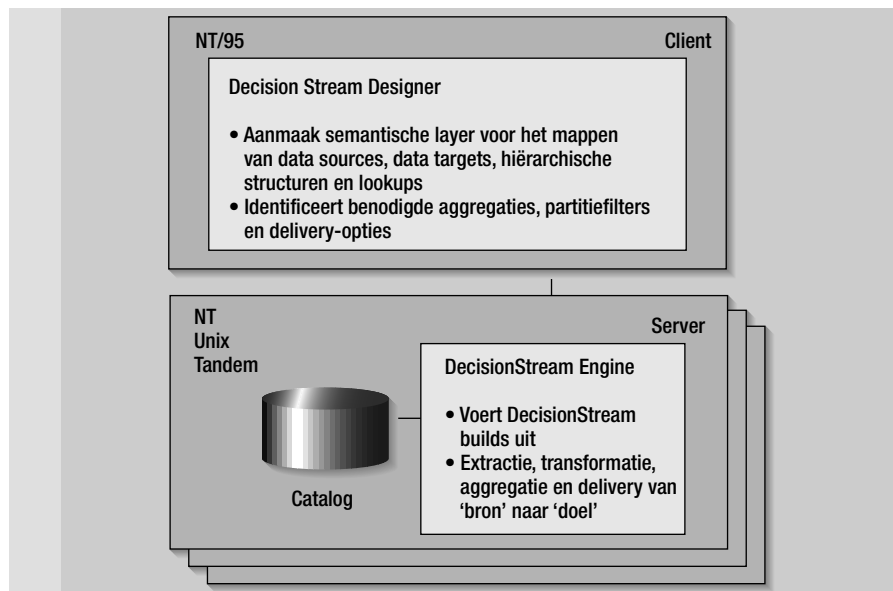
### DIRECT KUBUSSEN AANMAKEN

Een belangrijk voordeel voor het CAC van het standaardiseren op Cognos-producten voor analyse en rapportage is dat DecisionStream na het transformeren van de gegevens automatisch een PowerPlay Olap PowerCube kan aanmaken. Deze PowerCubes vormen de meerdimensionale kubussen die de zestig betrokken managers en coördinatoren in de diverse landen kunnen gebruiken voor het maken van analyses en het opstellen van rapporten.

Bovendien kan DecisionStream voor gebruik in Impromptu een catalog aanmaken op basis van de dimensie- en feitentabellen waarin de meetwaarden van de diverse dimensies vastliggen. Deze catalog ligt als een soort schil over het datawarehouse en wordt door Impromptu gebruikt om SQL-statements voor het vervaardigen



FIGUUR 3: HET DIMENSIONAL FRAMEWORK MAAKT HET MOGELIJK DIMENSIES ALS PRODUCT EN KLANT TE DEFINIËREN.



FIGUUR 4: EEN DATA STREAM VAN BRONSISTEEM TOT DOELSYSTEEM.

van rapporten samen te stellen.

Daarnaast wordt een Impromptu-rapport gegenereerd dat door eindgebruikers kan worden toegepast om van de geaggregeerde gegevens in de kubus door te kunnen prikken naar de onderliggende gege-

vens in het datawarehouse. Ook die functionaliteit speelde een belangrijke rol bij CAC's beslissing om over te stappen op een volwaardig datawarehouse. Bij het analyseren van de door het CAC geregistreerde meldingen is het hierdoor tevens

mogelijk de relevante individuele cases op te halen en nader te bekijken.

Een harde eis van de gebruikers was dat zij dagelijks inzage zouden kunnen hebben in de voor hen relevante cases. "Voor landenorganisaties is een goed zicht op de serviceprestatie van groot belang. Hebben ze een melding gekregen en die toegewezen aan een dealer, die vervolgens geen actie heeft ondernomen, dan is het belangrijk snel te weten waarom dat niet is gebeurd. Het kan een vergissing zijn, overmacht, de klant die een probleem heeft gemeld was al weer vertrokken, noem maar op. Willen we echter een hoog niveau van service kunnen garanderen, dan moeten de landenorganisaties -maar ook wijzelf- iedere dag kunnen beschikken over volledig bijgewerkte gegevens. En dat betekent dus iedere nacht een 'update' verzorgen."

Robbert Hoeffnagel  
is freelance journalist.

Vervolg van pagina 35.

modelleren als dimensie is slechts raadzaam als het aantal unieke nummers beperkt is. Voor deze specifieke vragen met betrekking tot detailgegevens kan als oplossing een aparte hybride-OLAP-kubus (Holap) worden gemodelleerd zonder aggregaties. Met als nadeel dat OLAP Connect hiervoor geen nette gebruikersinterface heeft bedacht en dat te allen tijde een meetwaarde wordt opgehaald. Ook is het jammer dat OLAP Connect in dat kader geen drill-through ondersteunt. De MDX-specificatie voorziet hierin wel. Voor het opvragen van gedetailleerde gegevens kan men toch maar beter gebruik blijven maken van de semantische laag.

## CONCLUSIE

Het product OLAP Connect is een waardevolle uitbreiding van Business Objects. De grote kracht is dat BO complexe analyses, toegang tot relationele én multidimensionale gegevensbronnen en geavanceerde reportingmogelijkheden combineert.

Hiermee komt een *single-point-of-entry* voor business intelligence in een organisatie een stap dichterbij. Voor de IT-organisatie achter het datawarehouse kan de tandem OLAP Connect/Analysis Services minder werk betekenen en een hogere

### Migratie van een semantische laag naar OLAP Connect brengt een forse inspanning met zich mee

kwaliteit. Complexe formules zijn in een handomdraai te maken. Responstijden kunnen eenvoudig worden gemeten en verbeterd. Beheer van de definities is sterk te vereenvoudigen.

Duidelijke minpunten zijn het zoeken naar detailgegevens, ondersteuning van vergelijkingen als kleiner-dan en groter-dan, drill-through, softlinks en de 'automatische' naamgeving van gegevens wanneer deze van de OLAP-server zijn opgehaald. Het is best verwarrend dat iedere

meetwaarde tweemaal aan de eindgebruiker wordt gepresenteerd in de lijst van variabelen. Bij het ophalen van de meetwaarde Store Sales is deze zichtbaar als 'Store Sales' en 'Store Sales (servervalue)'. Het is onduidelijk welke de gebruiker nu moet kiezen.

Migratie van een semantische laag naar OLAP Connect brengt een forse inspanning van zowel de IT- als de gebruikersorganisatie met zich mee. Zo'n veranderings-traject is alleen aan te bevelen als genoemde voordelen op dit moment een duidelijk struikelblok wegnemen voor de verdere professionalisering van de organisatie van managementinformatie. Een belangrijke indicatie kan daarbij zijn dat de organisatie op dit moment nog een gegevensgerichte focus op managementinformatie heeft en wil overgaan op een meer kengetalgeoriënteerde wijze van sturen.

Daan van Beek MSc (dvbeek@brocacef.nl) is coördinator van het data warehouse competence centre bij Brocacef, een groothandel in farmaceutische producten en geneesmiddelen.