



BIReady maakt veel handwerk bij datawarehouses overbodig

BI uit een doosje

Teus Molenaar

Naast een datawarehouse staat altijd wel een ploegje onderhoudsmonteurs klaar, poetslap en oliesluit bij de hand. Dat kost niet alleen veel geld, maar is ook tijdrovend, waardoor de motor eigenlijk steeds op drie bougies loopt, om in garagetermen te blijven. BIReady lost dit op door een groot deel van het handwerk te automatiseren. Een Nederlands bedrijf dat samen met IBM en Sogeti een BI-appliance heeft ontwikkeld.

Vooraf een citaat uit het rapport 'Pitfalls in data warehousing, 2007' van Gartner: "IT-organisaties moeten nooit vergeten dat het datawarehouse-model dynamisch is en nooit als af is te beschouwen. Het bedrijfsleven blijft veranderen. Als de datawarehouse-weergave hetzelfde blijft, dan vermindert de waarde van de data in het datawarehouse snel, en genereert het datawarehouse verkeerde/ineffectieve/ongeschikte antwoorden op query's van gebruikers. Een flexibeler en bedrijfsbewustere benadering van datawarehouses is noodzakelijk."

De toon is gezet, want bovenstaand citaat is precies waar het om draait bij BIReady: ervoor zorgen dat datawarehouses en datamarts 'bijblijven' vereist gewoonlijk een groep consultants die voortdurend sleutelen aan het model. Dat is kostbaar en kost bovendien veel tijd. Tegen de tijd dat de 'Algemene Bedrijfs Keuring' achter de rug is, zijn de bedrijfsregels meestal weer veranderd. Dit gegeven heeft Gertjan Vlug en Harm van der Lek ertoe gebracht het overgrote deel van de datawarehouse-onderhoudsbeurten te automatiseren. "Dat wil zeggen; Harm doet alles wat met de technologie heeft te maken, en ik doe de

organisatie en de marketing. Verder hebben we een deel van de productontwikkeling uitbesteed aan een groepje van vijf mensen", vertelt Vlug. In 2005 is gestart met het proces dat uiteindelijk BIReady is geworden: een datawarehouse-generator.

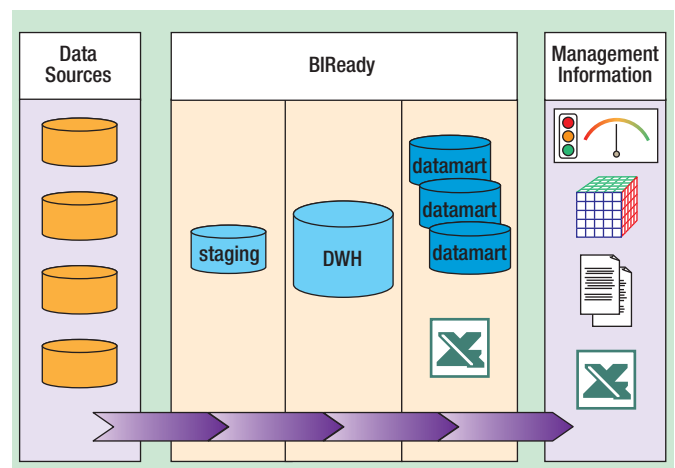
Corporate Information Factory

Eerst maar eens een duikje in de techniek. Vlug en Van der Lek menen dat fenomenen als translation, extraction, cleansing, datawarehouse keys, history, integration, design, create and load, en logging vrijwel altijd generieke patronen volgen. "Daarom zijn dergelijke processen ook te automatiseren", legt Van der Lek uit. Het tweetal meent dat elk datawarehouse in principe hetzelfde is, en dat het soort rapportages ook niet veel van elkaar verschilt bij alle organisaties. "Er is zoveel generiek dat je dat kunt automatiseren."

Bedrijven hebben behoefte aan managementinformatie uit de operationele systemen. Met transformatie en extractie komen die

BIReady, Inc. gepresenteerd tijdens Gartner BI Summit

Op 1 april 2008 werd tijdens de Gartner BI Summit in Chicago de Amerikaanse dochtermaatschappij van BIReady gepresenteerd. BIReady, Inc. houdt kantoor in New York en is geïncorporeerd als een Delaware Company. Het bedrijf gaat in nauwe samenwerking met technologie-partners en verscheidene system integrators en consultancy firma's de BIReady Data Warehouse Generator in de Amerikaanse en Canadese markt zetten.



Afbeelding 1: Architectuur van BIReady.



Harm van der Lek (l.) en Gertjan Vlug: met BIReady wordt saai werk geautomatiseerd.

gegevens in het datawarehouse en in datamarts die als subsets dienst doen. Die zijn uitgevoerd in een sterschema zodat ze snel bevroegbaar zijn. De een heeft genoeg aan een cockpit, de ander wil analyses doen en heeft dan meer detail nodig. Volgens Bill Inmon (Vlug noemt hem de 'godfather van datawarehousing') is zo'n model de Corporate Information Factory architectuur. "En die ondersteunen wij volledig", stelt Vlug.

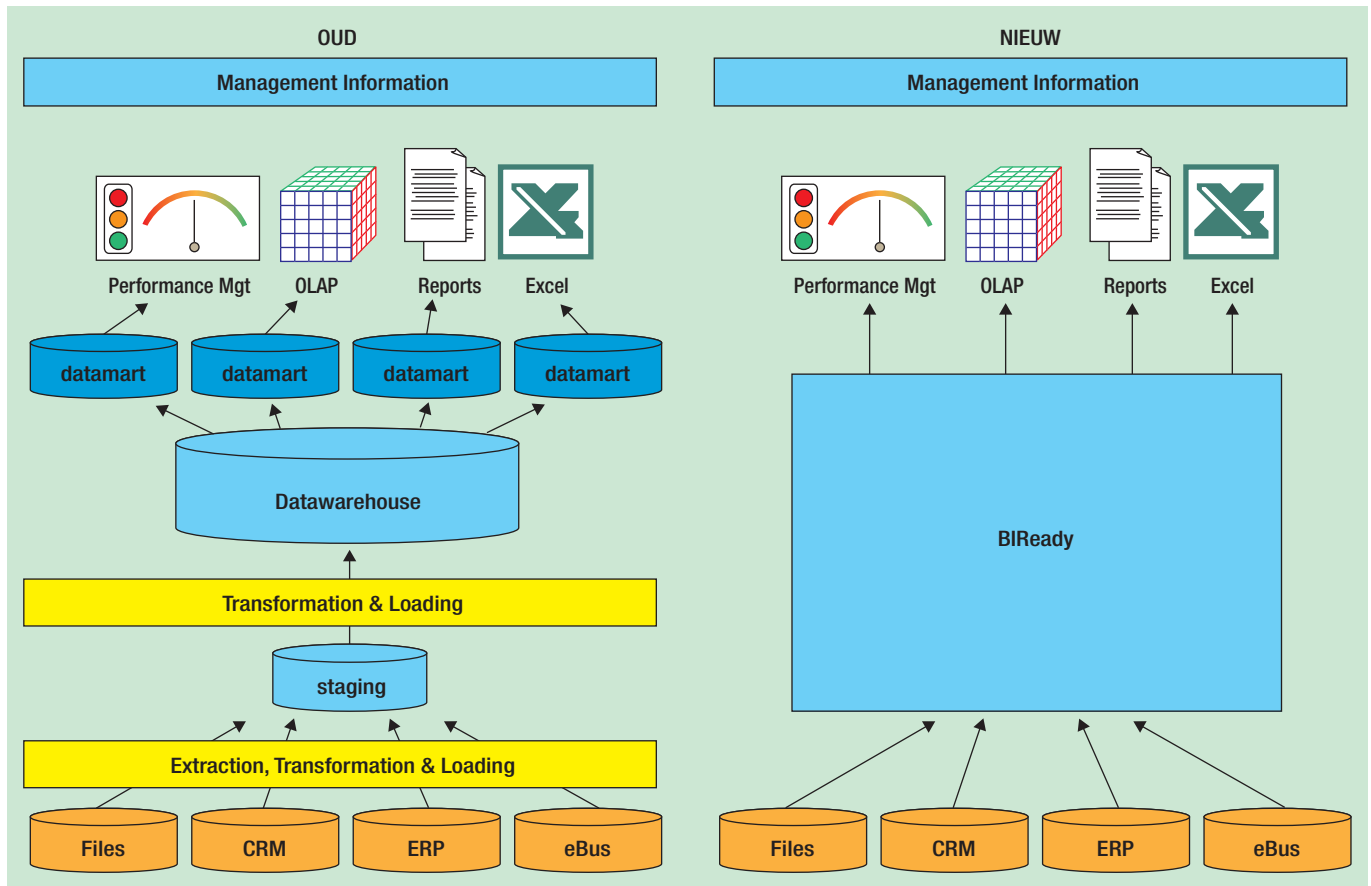
Documentatie en logging zijn twee stiefkindjes bij elk datawarehouseproject

In een datawarehouse heb je datawarehouse keys nodig; die worden met BIReady automatisch gegenereerd. Historische gegevens vormen de ruggengraat van het datawarehouse. Het kost veel tijd (en geld) om het datawarehouse te vullen. Vooral omdat vaak niet tevoren al bekend is welke gegevens je precies nodig hebt om de gewenste managementinformatie boven tafel te krijgen. "Wij gaan uit van de bedrijfsregels en van daaruit

genereer je automatisch de juiste historie", legt Vlug uit. Van der Lek vult aan: "Het gaat om de typen historie. Verzekeringsmaatschappijen bijvoorbeeld hebben niet genoeg aan simpele historie (waaraan je bijvoorbeeld kunt zien of en wanneer iemand is getrouwd en of dat gepaard ging met een verhuizing), maar willen ook een *terugwerkende kracht mutatie* kunnen verwerken. Nieuwe informatie kan de simpele historie dan veranderen. Je kunt een tabel herstellen – handmatig – maar dan ben je kwijt wat je eerst voor waar aannam. En verzeke-

BI op aanvraag

Samen met IBM en Sogeti heeft BIReady de Bibox ontwikkeld. Dit apparaat kent standaard ETL-functionaliteit, kan een datawarehouse en datamarts genereren, legt verbindingen met bronsystemen (dat kan ook een datawarehouse zijn), dient als een server en regelt opslag. Het lijkt op een BI-appliance, met dit verschil dat de Bibox niet is gebouwd voor één specifiek platform of een specifieke taak. Deze oplossing biedt BIReady ook aan volgens het SaaS-model, waarbij de 'intelligente doos' buiten een organisatie wordt gehost.



Afbeelding 2: Datawarehousing – traditionele oplossing versus de BIReady-oplossing.

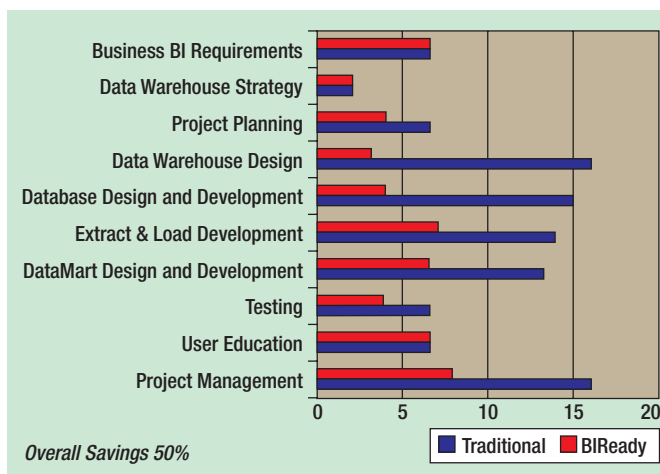
ringsmaatschappijen willen die kennis niet kwijt, want ze willen de rapportages kunnen verklaren die blijkbaar waren gebaseerd op verkeerde gegevens. Dat klinkt complex, dat is ook complex, maar het is wel generiek. Als je eenmaal weet hoe je dat moet doen, dan kun je dat kunstje herhalen. Wij hebben ervoor gekozen om, als je geen of weinig terugwerkende kracht mutaties hebt, in die tabel evenveel rijen op te nemen als wanneer je simpele historie zou hebben opgeslagen. Bij andere oplossingen wordt zeker de helft meer gereserveerd en dan krijg je op een

gegeven moment te maken met een ‘exploderend’ datawarehouse. Je moet dat op een economische manier regelen. Dit is een mooi voorbeeld van generieke functionaliteit. Dat kun je steeds weer op dezelfde manier doen.”

Oplossing ‘late arriving data’

Nog een voorbeeld van het nut van BIReady; hierbij gaat het om informatie die nog niet verwerkt kan zijn in het datawarehouse, omdat deze database gewoonlijk ‘s nachts zijn informatie uit alle onderliggende databases haalt. Stel dat iemand net een betaalrekening heeft geopend bij een bank, naar buiten loopt en geld uit de muur trekt. Zo’n mutatie kun je nog niet verwerken in het datawarehouse, omdat de details nog niet bekend zijn. “Dat is een heel lastig probleem”, zegt Van der Lek. “Wij lossen dat op door de transactie alvast in het datawarehouse te zetten en als de bankrekeninggegevens dan een of twee dagen later in het datawarehouse bekend raken, dan worden die gegevens automatisch aan elkaar gekoppeld. Ga dat maar eens handmatig doen; dat is een hell of a job.”

Vlug vertelt dat hij ook vaak heeft meegemaakt dat een dergelijke transactie geweigerd werd door het datawarehouse om te worden opgenomen, omdat die rekeninggegevens niet bekend zijn. “Dat kun je niet doen, want dan is het datawarehouse nooit compleet. Een transactie wordt namelijk maar één keer aangeboden.”



Afbeelding 3: Besparingen met BIReady.

Doorgewinterd

De mannen achter BIReady hebben hun sporen wel verdiend in het BI-veld. Gertjan Vlug heeft bij Cognos gewerkt en was een van de eerste medewerkers toen Business Objects (thans een onderdeel van SAP) in Nederland kwam. Hij was de eerste reseller ter wereld van Kalido, een datawarehouse-generator die een jaar of tien geleden door Shell is gebouwd; later is dit onderdeel verzelfstandigd en het is nog steeds actief op de markt. Het product belooft veel besparingen in tijd en kosten op datawarehousing. Het was wel erg duur en zeker niet weggelegd voor de kleinere bedrijven.

Om na te gaan of Kalido echt zo interessant is, belde hij met de Nederlandse datawarehouse-expert Harm van der Lek. "Hij heeft zich er goed in verdiept en gaf ook veel feedback aan Kalido. Daarna is hij aan de slag gegaan bij het UWV (toen nog GAK) waar verscheidene datawarehouses naast elkaar draaiden en elk zijn eigen werkteam had. Er werd van alles dubbel gedaan. Daar is toen iets generieks gebouwd, zodat je al die verschillende datawarehouseprojecten sneller kunt afronden. Dat draait er nog steeds", vertelt Vlug. Van der Lek is vooral bekend geworden door zijn boek 'Sterren en Dimensies' dat uitvoerig ingaat op het ontwerpen en onderhouden van datawarehouses. Daarbij maakt hij dankbaar gebruik van de zogenoemde sterschema-technieken. Van der Lek is een begenadigd spreker over dit onderwerp op tal van seminars.

Overigens zitten aan deze oplossing ook 'haken en ogen'. "Je zult een drempelwaarde moeten aangeven voor het aantal transacties dat je op deze manier wilt verwerken. Als het percentage records dat fout is in het datawarehouse over een bepaalde drempelwaarde – deze kun je zelf instellen – heengaat, dan stopt de load. Want je kunt je voorstellen dat een verkeerde load wordt aangeboden, dan wil je niet dat hij voor een miljoen rekeningnummers tal van attributen op onbekend zet, wachtend op de gegevens van de nieuwe rekeninghouders. Je kunt de drempels per sector instellen, want je kunt je voorstellen dat het voor de afdeling marketing minder nauw luistert dan bijvoorbeeld het datawarehouse, waaruit de accountants moeten putten."

Andere manier van denken

Documentatie en logging zijn twee stiefkindjes bij elk datawarehouseproject. "Dat wil niemand doen, want je kunt er niet mee scoren en het is heel saai werk", zeggen Vlug en Van der Lek. "Niettemin is het heel belangrijk dat het gebeurt. Met BIReady is ook dit saaie werk geautomatiseerd."

Waar het vooral om gaat, zo benadrukt hij, is een andere manier van denken. "Bij ons is het uitgangspunt het business model. Een organisatie heeft klanten, vestigingen in meerdere landen, producten, partners, medewerkers, enzovoort. En die hebben allemaal hun eigen attributen. Dat alles komt bij ons in de repository. Daar voeg je historie aan toe, per attribuut. En op basis van dit model genereert BIReady een staging model voor

het datawarehouse en de datamarts. Het datawarehouse is genormaliseerd, en de datamarts zijn ingericht volgens het sterschema om heel snel resultaten te kunnen boeken.

De structuur van het datawarehouse en de datamarts zijn verschillend. Het datawarehouse is geoptimaliseerd voor opslag, historie, ETL; en de datamarts zijn geoptimaliseerd voor query's. "We hebben voor die verscheidenheid gekozen, omdat sterschema's niet zo makkelijk zijn te veranderen – sterschema's zijn starschema's. Terwijl het datawarehouse wel flexibel moet zijn, want het moet veranderingen in het business model kunnen volgen. De business staging area moet je dan voeden met data uit de bronsystemen. En de bronbestanden, gevoed door de transactie-applicaties, ondersteunen het business model. Dus mag je aannemen dat het dan wel gaat lukken."

De datamarts zijn ingericht volgens het sterschema om snel resultaten te kunnen boeken

Van der Lek beschrijft het gemiddelde BI-project: "Consultants buigen zich over de vraag welke ETL- en *end user* tools ze nodig hebben. En dat is het, maar wat nog niet is doorgedrongen is de vraag of ze een datawarehouse-generator nodig hebben. Het is een nieuw concept. Als wij aankloppen, dan zeggen ze dat ze al een ETL-tool hebben, maar dat bieden wij niet. We hebben nog veel missiewerk te verzetten."

In Nederland gebeurt dat bijvoorbeeld bij IBM Information On Demand in juni. In Amerika pakte BIReady uit op de BI Summit van Gartner in Chicago.

Teus Molenaar is freelance journalist.

Hoge snelheden

Eén van de eerste gebruikers van BIReady is de Coram Group, een producent van badkamerproducten en -concepten. Het bedrijf bezit een centraal datawarehouse met datamarts op de regionale kantoren in Europa. De bestuurders willen graag informatie hebben van alle groepsonderdelen tegelijk. Randvoorwaarden: lage kosten, hoge snelheden (met Europese vestigingen en een in Japan), korte implementatietijd, flexibel met veranderingen in datamodellen.

BIReady bleek de uitkomst. BIReady ontwerpt, bouwt en vult het datawarehouse en de datamarts (zelfs onmiddellijk) louter en alleen op basis van het bedrijfsmodel van Coram. Elke verandering aan het model wordt automatisch doorgevoerd in het datawarehouse. Het definiëren van elementen en types (inclusief de complexe historische gegevens) gebeurt met één muisklik met de tool BIReadyManager. Omdat BIReady alleen veranderde data verwerkt, is de snelheid hoog.