

BIReady: ambitieuze concurrent voor Kalido klaar voor de markt

# Modelgedreven datawarehousing

Paul van der Linden

**Bijna vijftien jaar nadat Bill Inmon met 'Building the Datawarehouse' de noodzaak en onvermijdelijkheid van datawarehousing heeft duidelijk gemaakt, kan geconstateerd worden dat de meeste datawarehouses nog steeds handmatig gebouwd worden. Wie kijkt naar de verschillende componenten waaruit een datawarehouse-omgeving bestaat – zoals ETL, het datawarehouse zelf, de datamarts of de verschillende front-end toepassingen – moet concluderen dat dit slimmer kan.**

In een tijdperk gekenmerkt door de drive naar meer efficiency en kostenbesparing is een oplossing met minder handwerk en meer geautomatiseerde uitvoering dan ook bij voorbaat interessant. Het Nederlandse BIReady vormt zo'n aanpak. En biedt daarnaast een mooie oplossing voor het omgaan met historie – toch een van de hoofdtaken van een datawarehouse. Hoe interessant is dit product?

## Achtergrond

Een datawarehouse-omgeving kent vele componenten. In de eerste plaats is er de ETL-component die zich richt op het verzamelen, transformeren en laden van de data die het datawarehouse ontvangt uit de verschillende bronsystemen. De tweede component is het datawarehouse zelf. We hebben het dan over de 'grote bak met gegevens'. Deze database kent vaak een genormaliseerd datamodel. Vanuit het datawarehouse kunnen weer delen apart worden klaargezet. De rationale om weer een deel van de datawarehouse-data klaar af te zonderen is om optimale toegankelijkheid en betere prestaties mogelijk te maken.

**Het voordeel van de constructie is dat een organisatie klein en goedkoop kan beginnen**

Deze datamarts kunnen multidimensionele kubussen of gedenormaliseerde databases zijn (sterschema's). Verschillende front-end tools maken gebruik van de informatie in deze datamarts. Uit het Nationaal Datawarehouse Onderzoek (zie [www.dbm.nl](http://www.dbm.nl)) blijkt dat

de meest voorkomende toepassing nog steeds rapportage is, op afstand gevolgd door analyse (OLAP). Uiteraard zijn er nog andere componenten en datawarehouse-architecturen te bedenken. Waar het hier om gaat is dat we het over veel verschillende componenten hebben. Veelal worden deze componenten los van elkaar aangepakt en ingevuld. Dat leidt tot suboptimalisatie en meer handwerk (tussen de componenten) dan strikt gesproken noodzakelijk.

De oplossing ligt voor de hand: modelgedreven datawarehouse-ontwikkeling. Door centraal de vereisten op te slaan over de gewenste informatie (entiteiten, attributen, relaties) kan vanuit een business-model het datawarehouse (en een groot deel van de datawarehouse-omgeving) worden gegenereerd. Hiermee verschuift de focus van het technisch ontwikkelen van het datawarehouse naar het specificeren van het business-model. De software genereert op basis van het gedefinieerde business-model het onderliggende database-model. Een van de voordelen van deze aanpak is dat de kans op menselijke fouten (typefouten, vergeten om veranderingen door te voeren naar alle betrokken componenten) hierdoor afneemt.

Met haar modelgedreven datawarehouse-aanpak lijkt BIReady op hetgeen de Britse firma Kalido in 2001 in de markt heeft gezet. Het gelijknamige product is ontstaan uit productontwikkeling bij Shell en wordt inmiddels door met name grote, internationale bedrijven gebruikt. De principes waarop BIReady is gebouwd zijn eveneens in de praktijk ontstaan (project bij het UWV). De overeenkomst is nog groter aangezien BIReady ook voorziet in een uitgebreide manier om met historie van data om te gaan. Op sommige punten biedt BIReady zelfs meer: zo kan BIReady omgaan met gegevens met een kennisdatum die ligt voor een eerdere (reeds in het systeem bekende) kennisdatum. De vraag kan dus gesteld worden in hoeverre BIReady een alternatief is voor het duurdere Kalido.

Alvorens op die vraag in te gaan is het zinnig om nader stil te staan bij de twee eerder genoemde aspecten: modelgedreven datawarehouse-ontwikkeling en de afhandeling van historie. Dit zijn immers de unieke aspecten die beide producten bieden.

## Opzet

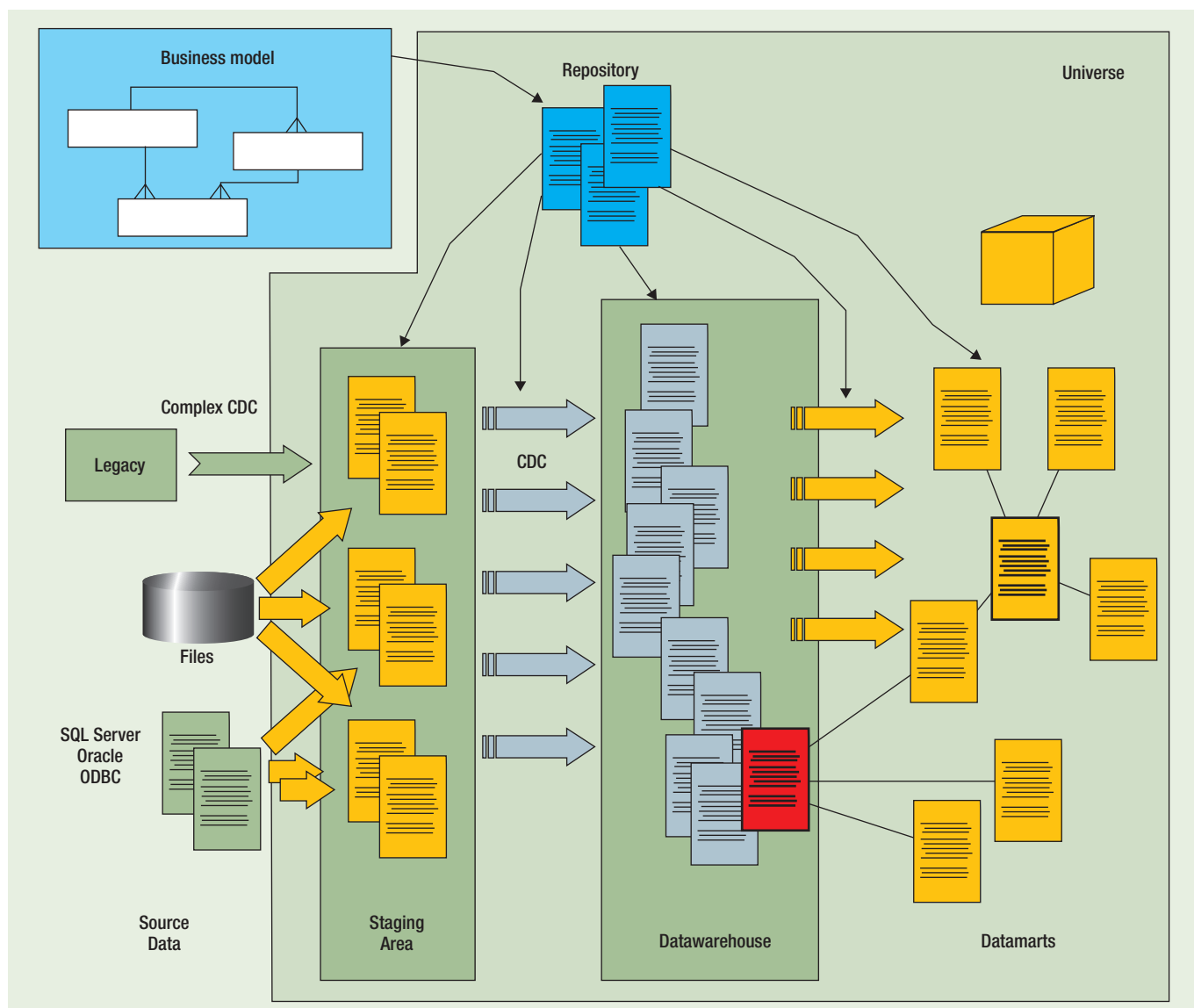
BIReady bestaat uit drie componenten: BIReady@Home, BIReadysMart en BIReadyGoGo. Hierbij vormt BIReady@Home

het product om de repository (of repository's) te maken en onderhouden en de staging area en het datawarehouse te genereren. Het is daarmee het centrale deel van het product. BIReadysMart zorgt voor het genereren van datamarts en BIReadyGoGo is een server-product gericht op het uitvoeren van procesinstructies. BIReady gebruikt minimaal drie verschillende omgevingen. Eén omgeving wordt gebruikt voor de repository en de user input. De tweede omgeving is voor het datawarehouse zelf. Per relationeel bronsysteem heb je vervolgens een aparte omgeving/mapping (omgeving 3 en verder). Een bron kan een tabel zijn, een view of een file (text delim, text fixed). Daarnaast kunnen via files en ODBC brondata worden ontsloten. Re-engineeren van bestaande databases wordt ondersteund. Per entiteit kan worden aangegeven welke attributen meegenomen moeten worden. Niet alle datatypes worden ondersteund (images of large text). Attributen met niet-ondersteunde datatypes worden genegeerd (niet meegenomen). Na afsluiten van de re-engineering kan een consistency check worden gedaan.

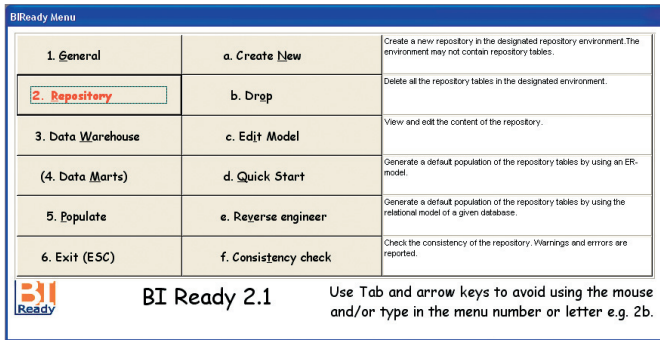
Via de optie Populate/Mappings kan vervolgens worden aangegeven waar de datawarehouse-attributen hun waarde vandaan moeten halen. Hierbij is een Try Defaults optie aanwezig. In de huidige versie wordt zelfs een default mapping gegenereerd voor de 'gere-engineerde' database. De volgende stap bestaat uit het creëren van het datawarehouse. Dit gebeurt vanuit de repository. Het vullen van het datawarehouse gebeurt met behulp van procesinstructies en BIReadyGoGo. Als er problemen optreden bij het vullen van het datawarehouse wordt aangegeven welke processen gefaald hebben en wat de reden hiervoor is.

## Modelgedreven

Niet voor alle attributen is dezelfde vastlegging van historie noodzakelijk. De eenvoudigste manier om historie vast te leggen is om alleen de actuele waarde te bewaren en de voorgaande waarde(n) te overschrijven. Dit is eigenlijk het niet vastleggen van historie. Ook is het mogelijk om een nieuwe waarde vast te leggen die met een bepaalde datum van kracht wordt. Deze ingangsd-



Afbeelding 1: Architectuur van BIReady.



Afbeelding 2: Menu: Repository.

tum kan in het verleden liggen, gelijk zijn aan de systeemdatum (dus nu ingaan) danwel in de toekomst ingaan. Het vasthouden van dit soort wijzigingen houdt technisch gezien in dat waarden niet worden overschreven, maar dat nieuwe records worden aangemaakt waarin de nieuwe waarde (en de bijbehorende ingangsdatum) wordt vastgelegd.

Gaat het om een zogenaamde 'temporal change' zoals BIReady onderscheidt, dan kan een wijziging leiden tot meerdere nieuwe records. In dit geval wordt de tijdelijke nieuwe waarde vastgelegd (nieuw record) en het weer ingaan van de oorspronkelijke waarde (nieuw record). Dit betekent dus dat BIReady werkt met aansluitende records, waarbij de begindatum van record X de einddatum van record X-1 bepaalt. Elke entiteit heeft standaard

dan ook een fromdate-veld en een todate-veld. Een derde tijdsattribuut (changedate) kan worden toegevoegd indien historie van de historie noodzakelijk is.

BIReady biedt de mogelijkheid om met zogenaamde historische blokken ('history blocks') te werken. Door met blokken te werken kan in het datawarehouse ruimte worden bespaard. Het uitgangspunt dat BIReady hierbij hanteert is de frequentie van verandering. Als voorbeeld wordt gegeven adres versus naamsverandering of geslacht. Aangenomen kan worden dat adres vaker zal veranderen dan naam of geslacht. Daarom zal adres in een ander historieblok zitten dan naam en geslacht. Deze indeling staat dus los van de gewenste historische variant.

## BIReady en Kalido

Kalido en BIReady zijn momenteel de enige producten die de mogelijkheid bieden om vanuit een model (repository) een datawarehouse-omgeving op te zetten, in te richten en te beheren. De verwachting is dat er meer van dit soort tools zullen komen. Deze ontwikkeling is vergelijkbaar met hetgeen in andere software-categorieën waargenomen kan worden. Een goed voorbeeld hiervan betreft de huidige OLAP-producten die kant en klaar gekocht kunnen worden. Niemand zal het meer in zijn hoofd halen om zelf OLAP-functionaliteit (drill down/up, slicen en dicen) te gaan bouwen. Dat de door BIReady en Kalido geboden functionaliteit zonder meer nuttig is staat buiten kijf. Wat grootschalige toepas-

---

sing van Kalido in de weg staat is de heftige aanschafprijs. BIReady lijkt die valkuil zorgvuldig te vermijden door een veel lagere instapprijs.

Hierbij moet wel worden opgemerkt dat Kalido (product en organisatie) een veel volwassener positie inneemt. De ontwikkeling van Kalido Dynamic Information Warehouse (DIW) – zoals het voluit heet – is terug te voeren naar midden jaren negentig. De marktintroductie is inmiddels vier jaar geleden. Organisaties zoals Unilever, ING, BP en Imperial Tobacco maken gebruik van Kalido. Het aantal werknemers en partners waarover Kalido beschikt is ook navenant groter. BIReady heeft daarentegen de ervaring van Harm van der Lek (een van de oprichters en de architect van het product) aan boord.

BIReady en Kalido bieden beide de mogelijkheid om wijzigingen door te voeren op een werkend datawarehouse. Kalido bewerkstelligt dit door een generieke aanpak waarbij alles intern wordt opgeslagen in slechts twee tabellen. Bij BIReady worden via de Adjust-functionaliteit wijzigingen doorgevoerd. Het datamodel bestaat hier uit meerdere tabellen. Hoe dan ook gaat het in beide gevallen om een constructie die zich voor gebruikers voordoet als een black box.

## Prijs en ondersteuning

De ambitie van BIReady is om eerst de Beneluxmarkt te bewerken en vervolgens snel de stap naar de Verenigde Staten te maken. Doel is om een wereldwijde software-leverancier te worden.

De prijs van BIReady@Home wordt bepaald door het aantal rollen in de repository (of de repository'). Een entiteit, de attributen, relaties en alle tijdsattributen nog apart, zijn allemaal rollen. BIReady wordt verkocht in zogenaamde packages bestaande uit 50 rollen. De prijs van een package is 50.000,- euro. Bij groei geldt een prijsstafel.

Het voordeel van een dergelijke constructie is dat een organisatie klein en goedkoop kan beginnen. Gezien de bedragen kan een en ander vanuit een beschikbaar projectbudget worden bekostigd.

Bij groei van de repository kunnen rollen worden bijgekocht. Wel moet bedacht worden dat in praktijk men snel over 50 rollen heen zal gaan. Een simpel datamodel telt toch al gauw 10 entiteiten. Met een gemiddelde van 10 attributen per entiteit zit je dan al over 100 rollen (historie en relaties meegerekend).

In praktijk moet dus waarschijnlijk eerder rekening worden gehouden met een beginprijs van 100.000,- tot 150.000,- euro (twee of drie packages).

De run-time applicatie (BIReadyGoGo) wordt meegeleverd voor de processen tussen staging area en het datawarehouse. Het genereren van datamarts gebeurt door middel van BIReadysMart. BIReadysMart is een optioneel onderdeel. Een van de redenen hiervoor is dat er een grote variëteit aan datamarts bestaat. Voorbeelden hiervan zijn PowerPlay-kubussen, MS Analysis Services cubes, BO Universes of ster- en snowflake-schema's in bijvoorbeeld een Oracle of DB2 database. Nog niet al deze mogelijkheden worden ondersteund. Momenteel wordt gewerkt

aan de mogelijkheid om vanuit het datawarehouse PowerPlay-kubussen te genereren. De relationele databases die momenteel worden ondersteund zijn MS Access, MS SQL Server en Oracle.

## Conclusie

Kijkend naar de functionaliteit die BIReady biedt is het zonder meer een interessant product te noemen. Modelgedreven datawarehouse-generatie vanuit een centrale repository biedt de voordelen van meer consistentie en een betere efficiency. Een uitgebreide en doordachte historie-afhandeling biedt daarnaast extra functionaliteit die aansluit bij de kerntaken van een datawarehouse. Dit zijn dan ook de onderscheidende factoren die BIReady te bieden heeft. Daarmee begeeft BIReady zich op het terrein waarop ook het Britse Kalido actief is, wat een soortgelijke functionaliteit biedt voor de Fortune 100-bedrijven.

## Per entiteit kan worden aangegeven welke attributen meegenomen moeten worden

Deze overeenkomst betekent dat BIReady en Kalido ook een belangrijk probleem delen: het probleem van de ETL-tool. Beide producten bieden beperkte ETL-functionaliteit, maar geven ook duidelijk aan geen ETL-tool te (willen) zijn. Hoewel dit waarschijnlijk een goede keuze is, zorgt het ook voor een belangrijke beperking. Je lost met BIReady een heleboel op, maar ETL blijft een belangrijk deel van je inspanningen (afhankelijk van de complexiteit van de omgeving).

Op een ander belangrijk punt onderscheidt BIReady zich van 'oudere broer' Kalido. Waar Kalido zich uitsluitend richt op grote, internationale ondernemingen en dus start met een instapprijs van 300.000,- dollar heeft BIReady zijn pijlen ook gericht op de rest van de markt. Een startprijs van 50.000,- euro moet het product binnen het bereik van de meeste (alle?) ondernemingen brengen. Sterker: met een dergelijke prijsstelling wordt het mogelijk om uit een projectbudget BIReady aan te schaffen hetgeen een langdurig en moeizaam beslissingstraject binnen de organisatie overbodig maakt. BIReady springt hiermee in het gat wat Kalido al jarenlang open heeft laten liggen.

Een interessant product? Zeker. Natuurlijk bevindt BIReady zich pas aan het begin van een hopelijk lange en vruchtbare onderneming die modelgedreven datawarehouse-ontwikkeling binnen het bereik van geïnteresseerde organisaties moet brengen. BIReady is klaar voor de markt. Nu maar hopen dat de markt ook klaar is voor BIReady.

### Paul van der Linden

Paul van der Linden (Paul.PFH.vanderLinden@AtosOrigin.com) is senior consultant Data Warehousing/BI bij Atos Origin en geeft leiding aan Data Warehousing Cost & Lifecycle Management (CLM).