

Business Activity Monitoring komt naderbij

Business Intelligence: drie werelden komen samen

Paul van der Linden

Business Intelligence (BI) is een bekende term, maar waar de grenzen ervan lopen is moeilijker om aan te geven. Als de definitie van BI luidt dat het gaat om sneller betere beslissingen te kunnen nemen bestrijkt BI maar liefst drie categorieën van software. Te weten: de ETL-tools, de databasemarkt en de markt voor frontend-tools. Bij de ETL-tools gaat het om het verzamelen en combineren van data. De databasemarkt betreft het vastleggen en beschikbaar stellen van deze data. En bij de frontend-tools staat het ontdekken en gebruiken centraal.

Trends in BI houdt derhalve in de signalering van ontwikkelingen in al deze drie categorieën en over deze categorieën heen. In dit artikel zal eerst aandacht worden besteed aan de ontwikkelingen op datawarehouse-gebied. Daarna komen ETL en de frontend-tools aan bod.

DATAWAREHOUSING

Bill Inmon presenteert zijn visie op datawarehousing onder de noemer Corporate Information Factory (CIF). Wie Inmon door de jaren heen heeft gevolgd, kan concluderen dat er in de loop van de tijd steeds meer databasecomponenten zijn gedefinieerd. Voor wie het nog niet wist: Bill Inmon is degene geweest die duidelijk heeft gemaakt dat er naast de operationele (ook wel: transactionele) omgeving nog een andere omgeving nodig is. Deze heet:

de datawarehouse-omgeving (en nu dus: CIF).

In de OLTP-omgeving staan de verschillende applicaties centraal. Het gaat hierbij bijvoorbeeld om inkoop, verkoop, marketing en personeel. Daarnaast is een omgeving nodig waarin de data centraal staat: de datawarehouse-omgeving.

Belangrijkste component hiervan is het datawarehouse zelf. Maar ook de operationele datastore (ODS), de ETL-component, de datamarts en de DSS-applicaties hebben hierin hun plaats. De CIF, die kan worden opgevat als de datawarehouse-omgeving zoals we die nu kennen, bevat naast de al eerder genoemde componenten nog achteraf toegevoegde componenten zoals de exploration-warehouses en de datamining-warehouses. ETL is inmiddels door Inmon herdoopt in: integratie & transformatie (I&T).

Inmons tegenpool Ralph Kimball geeft in zijn column in Intelligent Enterprise

van 26 juli 2002 aan dat het tijd is om iets te doen wat hij al een hele tijd niet heeft gedaan: de grondbeginselen uitleggen van datawarehousing¹. Waarom was het ook alweer nodig en waarom deden we bepaalde dingen? Het datawarehouse als fenomeen is terechtgekomen in een positie waarin het duidelijk is dat de geldende theorie niet meer voldoet, maar waarin het nog onduidelijk is hoe het dan wel zou moeten. Inmon probeert het op te lossen door steeds meer gegevensverzamelingen toe te voegen aan zijn CIF-plaatje. Kimball gaat terug naar de basis en komt straks wellicht met nieuwe antwoorden op oude vragen. Want één ding is duidelijk: de omstandigheden waarin het datawarehouse dienst doet zijn inmiddels ingrijpend veranderd. Dat is niet in elke organisatie in dezelfde mate voelbaar en zichtbaar, maar het valt niet meer te ontkennen. Het tempo waarmee informatie moet worden opgehoest is alleen maar toegenomen. Real time en zero latency zijn de termen



die hierbij horen. Aan de andere kant neemt de hoeveelheid te verwerken data alleen maar toe. Het aandeel van de gestructureerde data (die makkelijk kan worden weergegeven in rijen en kolommen) neemt procentueel gezien af. Het aandeel van de ongestructureerde data, zoals e-mails, is alleen maar groeiende. De te beantwoorden vraag is derhalve hoe om te gaan met explosief groeiende datavolumes, die steeds sneller op ons af komen en steeds sneller beschikbaar moeten zijn, en die steeds minder goed passen in onze relationele databases?

FOCUSVERSCHUIVING

Na de applicatiecentrische focus zoals de OLTP-omgeving die kent, en de daarop volgende datacentrische focus van het datawarehouse is er nu sprake van een verdere verschuiving. De focus is aan het verschuiven naar de processen. Het integraal kunnen aansturen en meten van activiteiten staat nu centraal. Gartner noemt dit 'business activity monitoring' (BAM). Het werken met key performance indicatoren (kpi's) en een balanced scorecard (bsc) zijn hier voorbeelden van. Of eenvoudiger gezegd: na een periode waarin we ons hebben blind gestaard op de data is het nu tijd om er daadwerkelijk mee aan de slag te gaan en actie te ondernemen; toe te juichen dus als een teken van verdere volwassenheid van het datawarehouse.

DATAWAREHOUSING NIEUWE STIJL

Terwijl dat de mainstream databaseleveranciers IBM, Oracle en Microsoft het merendeel van de datawarehouse-databases verzorgen, verschijnen er ook nieuwe spelers aan de horizon. Kalido heeft met Dynamic Information Warehouse de voortdurende veranderende omgeving waarin datawarehouses moeten opereren als uitgangspunt genomen. Het wordt dan ook gepresenteerd als de nieuwe manier om datawarehouses te bouwen. Kalido is met name een goede oplossing als het erom

gaat om verscheidene, reeds bestaande datawarehouses te integreren. Dit blijkt ook uit de boodschap zoals Kalido Group die uitdraagt: 'help enterprises manage complex change such as mergers, acquisitions, reorganisations and globalisation' - zeer actueel in het huidige tijdsgewricht. De oplossing die Kalido levert bestaat uit: Information Server (datareplicatie), Enterprise Modeller (KPI en CSF), Aggregation Services (data-charts en cubes) en Distributed Services (import/export).

ETL

Informatica is wellicht de bekendste speler op het gebied van de extractie, transformatie en het laden van data. Ze heeft haar naam gevestigd met producten als PowerMart en PowerCenter. De ETL-markt die een zeer winstgevend onderdeel van de softwaremarkt vormde en waarin gigantische prijzen gevraagd (en betaald) werden is inmiddels zo geërodeerd dat een koersverandering noodzakelijk is geworden. De introductie van Data Transformation Services (DTS) door Microsoft in 1999 (als onderdeel van het databaseproduct SQL Server) heeft gezorgd voor een neerwaartse druk op de prijs voor ETL-tools; waarmee de winstgevendheid van ETL-software als op zichzelf staand product is afgenomen.

ETL en datawarehouses vormen de backend-componenten van business intelligence-oplossingen. In de praktijk blijken bedrijven nog steeds eerder bereid om geld uit te geven aan flashy OLAP-producten dan aan deze noodzakelijke backoffice-software. Ook hierin kan geconstateerd worden dat Microsofts aanpak om business intelligence gebundeld aan te bieden (backoffice- en frontoffice-componenten) een trend heeft gezet.

INTEGRATIE VAN ANALYTICS

Dit is dan ook een van de ontwikkelingen die momenteel kunnen worden waargenomen in de ETL-markt. Vroegere ETL-only

leveranciers completeren hun softwareaanbod door analysesoftware toe te voegen. Informatica anno 2002 profileert zich derhalve als 'leading provider of data integration and analytics software'. Door samenwerking met het bedrijf WebMethods wordt vormgegeven aan Business Activity Monitoring (BAM). Gaurav Dhillon, ceo van Informatica, merkt op dat BAM het natuurlijke gevolg is van het combineren van real-time applicatie-integratie, zero latency data-integratie en business analytics. Het gaat hierbij om het vermogen analyse toe te passen op informatie die komt uit verscheidene applicaties, databases en de bedrijfsprocessen die de onderlinge relaties tussen deze resources behe-

ETL en datawarehouses vormen de backend- componenten van BI-oplossingen

ren. PowerMart en PowerCenter zijn derhalve niet meer de producten die bij Informatica in de schijnwerpers staan. Dat is inmiddels voorbehouden aan de Applications Analytic Software Family. Hiermee wordt een reeks software aangegeleid die zich richt op customer relationships, web channel analyses, business operations, strategic sourcing en de supply chain.

ETL & EAI

Een tweede ontwikkeling die kan worden waargenomen is van meer technische aard. ETL-software heeft zich altijd gericht op het batchgewijs overzetten van grote hoeveelheden data. Deze manier om data te transporteren is niet altijd meer voldoende. Door toepassingen zoals customer relationship management (CRM) of door e-business toepassingen worden eisen gesteld ten aanzien van de tijdigheid waarmee informatie aanwezig dient te zijn. En steeds meer gaat het hierbij om real time, zero latency oplossingen (zoals

Dhillon al stelde). Informatie dient beschikbaar te zijn op het moment dat ze nodig is, niet op het moment dat de batch-verwerking toevallig is afgerond.

De tijdsperiode die beschikbaar is om batches uit te voeren (het batchwindow) is in de loop van de tijd alleen maar kleiner geworden, terwijl de hoeveelheid informatie die moet worden bewerkt alleen maar toeneemt. Ook vanuit deze achtergrond is duidelijk dat een batch-only oplossing op de lange duur niet voldoende is. Hieruit is een huwelijk ontstaan tussen de batchgewijze ETL-oplossingen en de messaging zoals bekend vanuit de Enterprise Application Integration (EAI) wereld. Het bedrijf Acta profileert zich als 'het eerste en enige real time data-integratieplatform'. De combinatie van ETL en EAI belooft in ieder geval veel goeds. ETL kan veel data aan, maar niet real time. EAI levert het omgekeerde plaatje zodat een combinatie elkaars zwakke punten verwijdert. Wat real time moet, kan via de EAI-kant worden gerealiseerd. De rest kan op de standaard ETL-manier blijven plaatsvinden.

Overigens is Acta 'slachtoffer' geworden van de eerste trend. Business Objects (BO) heeft Acta overgenomen, iets wat zijn beslag zal krijgen in het derde kwartaal van 2002. Volgens BO betekent dit dat ze nu een 'comprehensive enterprise analytic platform' biedt. Hiermee wordt bedoeld op de combinatie van BI en data-integratie. De overname die BO doet staat niet op zichzelf. Informatica's beweging richting analytics, leidend tot de beëindiging van een licentieovereenkomst die Informatica had met BO, heeft de laatste geïnspireerd tot een tegengestelde beweging richting data-integratie. Informatica en WebMethods, BO en Acta zijn dus twee voorbeelden van de convergerende werelden van data-integratie en BI.

FRONTEND-TOOLS

De webificatie van het datawarehouse betekent dat de data die in het datawarehouse ligt opgesloten, bereikbaar wordt voor grote groepen gebruikers.

Leverancier :	Positie :	Aandeel :
Hyperion Solutions	1	21,3
Microsoft	1	21,3
Cognos	3	12,1
Oracle	4	7,1
MicroStrategy	5	6,9
Business Objects	6	6,6
Applix	7	2,5
Cartesis/PwC	8	2,5
Comshare	9	2,4
IBM (DB2 OLAP Server)	10	2,1
Adaytum	11	1,5
Brio Technology	12	1,3
SAS Institute	13	1,2
Crystal Decisions	14	1,0
CA	15	0,3

TABEL 1: TOP 15 OLAP-TOOLS

Softwareleveranciers hebben in het verleden webversies ontwikkeld van hun client/server-analysesoftware. Denk hierbij aan WebIntelligence van BO of aan PowerPlay Web van Cognos. Deze webversies zijn tot op heden echter qua functionaliteit achtergebleven op de c/s-versies waarop ze zijn geënt. Leveranciers van frontend-software zoals BO zijn nu bezig om deze achterstand in te halen. Hiermee wordt webontsluiting van het datawarehouse meer dan voorheen een volwaardig alternatief en komt 'BI any place any time' dichterbij.

Frontend-tools komen voort vanuit een achtergrond van ad hoc query's en rapporten op relationele databases. Vervolgens zijn ROLAP en vervolgens MOLAP toegevoegd aan het databaseproduct. Momenteel zitten OLAP, ETL en datamining als onderdeel bij elk databaseproduct van de leidende databasleveranciers (IBM, Microsoft en Oracle).

OLAP

De OLAP-tools worden steeds grafischer en intuïtiever. Voorbeelden hiervan zijn Cognos NovaView dat beschikt over een uitgebreide set van mathematische, statistische, financiële en multidimensionale functies. Met behulp van point en click kan de gebruiker probleemloos analyses

verrichten. Ook Microsofts Data Analyzer, verkregen door de overname van het Israëlische bedrijf Maximal Innovative Intelligence, biedt een vlotte toegang tot en omgang met de data. Voor al deze producten geldt dat data in verschillende formaten zoals in tabelformaat, staafdiagrammen, pie charts, enzovoort kan worden weergegeven en dat wisselen van formaat probleemloos gaat.

Voor 2001 zien de posities en markt-aandelen (%) van de verschillende leve-

Datamining is altijd als een grote belofte gepresenteerd.

ranciers er volgens Nigel Pendse, auteur van The OLAP Report, er als volgt uit (op basis van gegenereerde omzet). De gedeelde eerste plaats wordt in 2001 ingenomen door een sterk stijgend Microsoft en een terugvallend Hyperion wiens Essbase-verkopen te wensen overlaten.

DATAMINING

Datamining is altijd als een grote belofte gepresenteerd. Maar behalve de echt

		Markt		
		ETL	Databases	Frontend-tools
Leveranciers	ETL	X		-> X
	Databases	X <-	X	-> X
	Frontend-tools	X <-		X

TABEL 2: LEVERANCIERS BEGEVEN ZICH OP MEERDERE MARKTEN WAARDOOR TRADITIONELE CATEGORIEËN NIET MEER GELDEN.

Ontwikkelingen in datawarehousing en BI
• BI-processen en hun onderlinge samenhang staan centraal (Business Activity Monitoring)
• ETL-leveranciers completeren hun softwareaanbod met analysesoftware
• Analyse- en ETL-faciliteiten van databaseproducten worden verder uitgebreid
• FET-leveranciers integreren ETL-functionaliteit in hun softwareoplossingen
• Samengaan van batchgewijze ETL-oplossingen met messaging-oplossingen uit de EAI-wereld
• Laagdrempelige datamining wint terrein
• OLAP-software steeds krachtiger en intuïtiever
• Grotere gebruikersgroepen in organisaties aan de slag met OLAP en/of datamining
• BI fact gap vereist coördinatie en afstemming in een BI-functie en BI-architectuur

TABEL 3: ONTWIKKELINGEN IN DATAWAREHOUSING EN BI.

grote bedrijven die de resources hebben om een eigen datamining-groep op te zetten heeft het deze belofte niet ingelost. Datamining is echter aan een tweede leven begonnen door nu laagdrempelig bij organisaties binnen te druppelen, bijvoorbeeld als meegeleverde software bij een aangeschafte database. Zo bouwt MS SQL Server 2000 met de nieuwe Data Analysis en Data Mining-tools verder op de koers die ze met versie 7 van het product zijn ingeslagen (OLAP Services, English Query en Unicode-ondersteuning). Decision Tree en Neural Network vormen het datamining-deel ervan. Vergeleken met SPSS Clementine, volgens de Kdnuggets website momenteel het populairste datamining-pakket, en startend met een prijs van \$ 55.000 per server (1 client) is de prijsdrempel ook veel lager. Aan de bovenkant van de datamining-markt biedt SAS Enterprise Miner een volledig pakket voor de ervaren dataminers (niet voor leken). In versie 4.1 is een aparte datamining-database (DMDB) toegevoegd. De data die hierin wordt opgeslagen is geoptimaliseerd voor datamining-gebruik. Zo zijn bijvoorbeeld variantie- en covariantiegegevens hierin al opgeslagen. Het lijkt

erop dat deze aanpak zal worden nagevolgd door leveranciers als Oracle, IBM en Microsoft.

ORGANISATIE VAN BI

Bureaus als Gartner hebben voorspeld dat organisaties steeds meer te maken zullen krijgen met een BI fact gap. Deze situatie kenmerkt zich door enerzijds te veel te verwerken informatie en anderzijds te weinig mensen die zich met deze taken kunnen bezighouden. Om dit probleem te adresseren kunnen verschillende oplossingen worden ingezet. De eerste is om de IT-afdeling(en) te ontlasten door verschillende BI-taken bij gebruikers neer te leggen. Dit betekent uiteraard wel dat deze gebruikers voorbereid (opgeleid) moeten zijn voor de taken die op hen afkomen. Een andere oplossing is van meer strategische aard. Alle BI-aspecten binnen het bedrijf dienen te vallen binnen een in te richten BI-functie. Deze BI-functie kan op verschillende manieren worden ingevuld, maar duidelijk is dat op verschillende niveaus (organisatorisch, software, hardware) coördinatie en afstemming vereist

zijn. Alleen op deze manier kan de databerg afgegraven worden en kan voordeel worden gehaald uit de investeringen die gedaan zijn in BI.

Deze ontwikkeling staat niet op zichzelf en heeft te maken met de algehele economische situatie. Waar het voorheen relatief makkelijk was om nieuwe investeringen in soft- en hardware te doen, ligt nu meer dan voorheen de nadruk op het beter laten samenwerken van hetgeen al aanwezig is binnen de organisatiemuren. Met een duur woord: architectuur. Dit geldt voor alle processen, data en applicaties, maar ook specifiek voor business intelligence (BI-architectuur). Deze trend is inmiddels opgepakt door softwareleveranciers, consultants en de organisaties waarin het allemaal moet gebeuren.

TOENADERING

Meer dan ooit gaat het om het bieden van complete oplossingen. Leveranciers van ETL-tools, databases en frontend-tools zullen hierdoor steeds meer op elkaars terrein komen. De werelden van data-integratie en analytics komen in hoog tempo samen. OLAP en datamining-software worden steeds krachtiger, maar ook visueler en laagdrempeliger. Hiermee worden ook grotere groepen gebruikers binnen organisaties geconfronteerd. Business activity monitoring wordt daardoor haalbaar, zeker wanneer vanuit een architecturele aanpak (bij voorkeur onder regie van een BI-functie) aandacht wordt besteed aan implementatie en gebruik.

Noten

1. 'An Engineer's View. It's worthwhile to remind ourselves why we built data warehouses the way we do', Ralph Kimball in Intelligent Enterprise, 26 juli 2002.

Paul van der Linden is senior consultant business intelligence bij Atos Origin. Paul.PFH.vanderLinden@AtosOrigin.com