



Het 15 jaar oude fundament begint scheuren te vertonen

Trends 2007 in BI en datawarehousing

Paul van der Linden

Business Intelligence is hot. Het staat vaak bovenaan of tenminste in de top-3 op het lijstje van aandachtspunten van de CIO. Tegelijkertijd is de traditionele visie op BI en datawarehousing aan het veranderen. De contouren van nieuwe BI en datawarehousing tekenen zich reeds af.

Zoals op het congres Datawarehousing & BI (Noordwijkerhout, 2 november) al is gemeld wordt de huidige BI-markt gekenmerkt door drie categorieën van spelers. Op de eerste plaats zijn er de grote BI-vendors die zowel een BI- als een datawarehousing-oplossing bieden. De bekendste drie leveranciers die in deze *pure play* categorie vallen zijn Business Objects, SAS en Cognos (in volgorde van omzet¹). Deze categorie bedraagt in totaal iets meer dan 43 procent van de markt (nieuwe en onderhoudslicenties).

Groei en innovatie

De grootste groei valt echter niet in deze categorie, maar in die van de enterprise application vendors. Denk daarbij aan Microsoft, Oracle en SAP. Qua omzet zijn deze leveranciers vele malen groter dan de pure play leveranciers. Belangrijk is dat deze leveranciers een hele andere visie op BI hebben. Voor de enterprise application vendors is BI een onderdeel van de processen die plaatsvinden binnen een organisatie. Het is dus niet een aparte activiteit die je op een apart platform met een aparte software-suite invult. Dit zijn ook de uitdagingen die pure play vendors als BO moeten oplossen (zie een eerder interview met BO's Bernard Liautaud in DB/M²).

De innovatie in BI komt van een derde categorie: die van de kleinere, gespecialiseerde BI-leveranciers. Hierbij kun je denken aan Spotfire, QlikTech en het Nederlandse Anago. Zij zijn beter dan de grote, gevestigde leveranciers in staat en bereid om gebruik te maken van technologische vernieuwingen zoals 64-bit architectuur, waarmee meer data ingelezen kunnen worden. Voeg daarbij een goede intuïtieve interface en je hebt opeens een stuk software waarmee de business daadwerkelijk zelf aan de slag kan.

Tijdens het congres werden drie scenario's toegelicht en aan de zaal gevraagd om in te schatten welk scenario in 2010 de meest voorkomende zal zijn. Met overgrote meerderheid werd gekozen voor het scenario waarin BI en datawarehousing worden ingevuld met behulp van een enterprise application vendor, met aanvulling door kleine specifieke BI-vendors voor specifieke toepassingen zoals bijvoorbeeld *deep analysis*. Dat enterprise applicatie-software (in 2010!) voldoende BI-functionaliteit biedt werd als minder waarschijnlijk gezien. Hetzelfde gold voor het overleven van de huidige dominante opzet waarin we twee verschillende omgevingen hebben (transactueel en analytisch). De grote pure play vendors lijken dus in de verdrukking te komen en gemengd

'Oud' (2006)	Nieuw (2007)
Inmon-opstelling met veel dataopslagstations, fysieke kubussen en datamarts	Geen of alleen virtuele kubussen (64-bit architectuur) Geen datamarts (werken op DWH)
Weinig data (Gigabytes)	Veel data (Terabytes)
BI = gestructureerde data	BI = ook ongestructureerde data
BI gericht op managers en analisten	BI voor de totale organisatie (information democracy)
Vertraging in beschikbaarheid BI-data	Real-time data
Hardcoded functionaliteit (monolithisch)	Functionaliteit via webservices (SOA)
BI als aparte activiteit (apart platform, aparte software-suite)	BI als onderdeel (aspect) van operationele processen
BI gaat over het opslaan en ter beschikking stellen van data	BI gaat over het gebruik van informatie ten behoeve van betere en snellere besluitvorming
Voornamelijk technische invalshoek	Meer aandacht voor de businessaspecten van BI

te worden door aan de ene kant de macht en kracht van de enterprise application vendors en aan de andere kant de snelle, technologische vernieuwing van de nieuwkomers.

Hoe het zover is gekomen

De grote pure play vendors hebben de afgelopen periode de ene na de andere partij opgekocht om tot een volledig aanbod van BI en datawarehousing te komen. Hiermee hebben ze gehoor gegeven aan de wens van bedrijven die het aantal gesprekspartners willen terugdringen. Maar ook de enterprise application vendors hebben die wens gehoord. Naarmate zij meer BI en datawarehousing bieden beschikken organisaties ineens over twee platformen met grotendeels overlappende functionaliteit. Optimalisatie zal waarschijnlijk ten nadele van de grote pure play vendor uitvallen. Zeker als ontbrekende, specifieke add-on functionaliteit gecombineerd kan worden met de software van de enterprise application vendors. Zie hier de levensgrote uitdaging voor de grote pure play vendors. Natuurlijk zijn ze morgen niet uit het beeld verdwenen. Maar de strijd om het leiderschap in BI en datawarehousing is verre van gestreden.

De overnamegolf betekent overigens niet dat het aantal spelers in de BI-markt is afgenomen. In tegendeel, er komen nog steeds nieuwe spelers bij – sommige vanuit heel andere delen van de IT (search engines, knowledge management, workflow management). Ook hier is dus geen sprake van een volwassen, gestabiliseerde markt, maar eerder van een dynamische en gevarieerde markt.

Ontwikkelingen in BI en datawarehousing

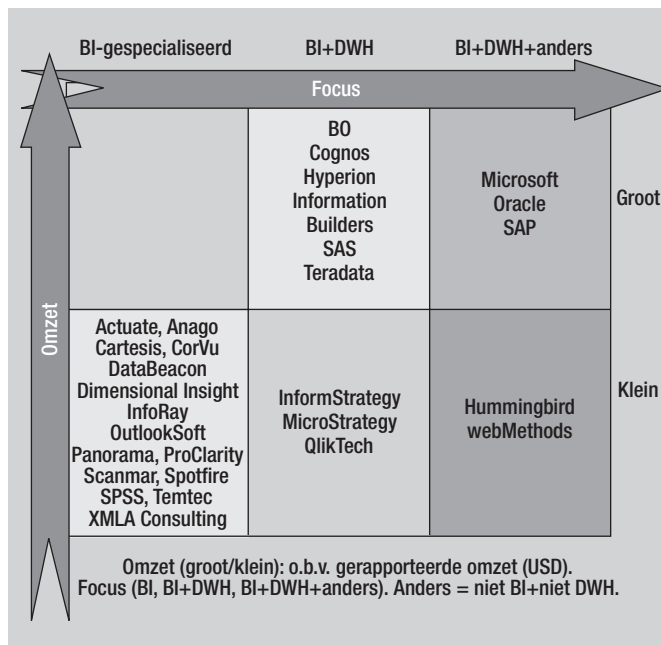
Een tiental ontwikkelingen zal achtereenvolgens besproken worden: 1. Leveranciers completeren hun aanbod; 2. 'Performance management als volgende stap in BI'; 3. Exponentiële datagroei; 4. Snellere beschikbaarheid data gewenst; 5. Information democracy; 6. Technologie, van 32-bit naar 64-bit architectuur; 7. Search engines; 8. DWH appliances; 9. Masterdata en meta-data; 10. Service Oriented Architecture.

1. Leveranciers completeren hun aanbod.

De grote partijen in BI en datawarehousing blijven doorgaan met het opslokken van kleinere partijen om hun aanbod verder te completeren. Voorbeelden hiervan zijn de overname van ProClarity door Microsoft, de overname van Siebel door Oracle en de overname van Crystal door Business Objects. De top-10 vendors in BI hebben volgens marktresearch bureau IDC dan ook de afgelopen jaren in marktaandeel gewonnen. Van een gezamenlijke 57,9 procent in 2003 is het marktaandeel voor de top-10 inmiddels opgelopen tot 62,3 procent in 2005. Deze beweging zal de komende jaren doorzetten. Of het een winnende strategie zal blijken te zijn is echter nog maar de vraag.

2. 'Performance Management als volgende stap in BI'

Performance management wordt vaak voorgesteld als de natuurlijke opvolger van BI. Dat is het niet. Het is een toepassing van ver-



Afbeelding 1: Omzet en Focus van de drie categorieën.

schillende BI-componenten (rapportage, analyse, datamining). Interessant, maar slechts een toepassing van BI en niet de volgende volwassenheidsfase ervan. Dat wordt al duidelijk als je bedenkt dat er ook andere toepassingen als customer relationship management (CRM) en compliance zijn.

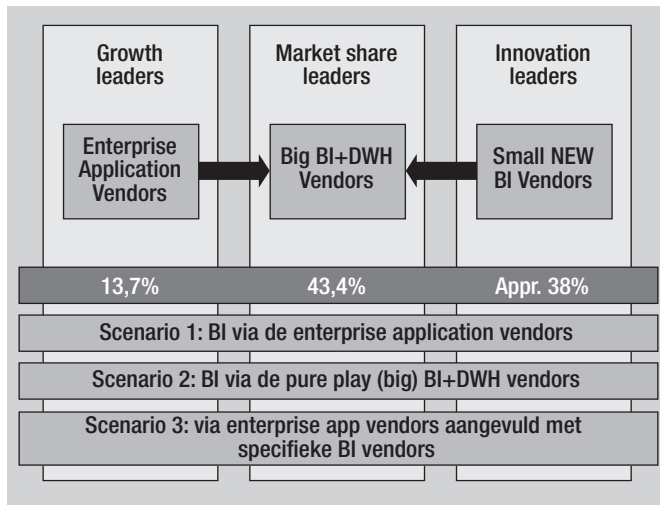
Met performance management wordt het proces aangeduid om de hele organisatie te laten werken aan dezelfde doelstellingen. Dit behelst de vertaling van de corporate strategie naar de verschillende afdelingsstrategieën. Strategy maps, key performance indicatoren (KPI's) en scorecards kunnen onderdeel zijn van een performance management aanpak. Voeding met de behaalde resultaten (actuals) vindt vaak plaats vanuit een datawarehouse. Gartner schat in dat momenteel minder dan 5 procent van de daarvoor in aanmerking komende organisaties gebruik maakt van een performance management suite³.

3. Exponentiële datagroei.

De hoeveelheid unieke informatie wordt op dit moment geschat op vijf Exabytes. In de laatste 30 jaar is meer informatie geproduceerd dan in de voorafgaande 5000 jaar⁴. Ook in termen van databases leidt dit tot extra belasting. Volgens de Winter corporation is de grootste commerciële database in de afgelopen drie jaar verdrievoudigd tot meer dan 100 Terabyte. In de afgelopen twee jaar is het aantal records en rijen van de allergrootste databases toegenomen tot 2,8 triljoen rijen (2005). Dit is een verviervoudiging ten opzichte van 2003⁵. Voor de meeste organisaties zullen minder grote aantallen gelden, maar dat de toename van de hoeveelheid data exponentieel verloopt is onmiskenbaar.

4. Snellere beschikbaarheid data gewenst.

Door ontwikkelingen als RFID (radio frequency identification),



Afbeelding 2: Drie scenario's.

performance management, Customer Relationship Management (CRM) en compliance-vereisten worden steeds meer gegevens gegenereerd en opgeslagen. Tel daarbij op de zogenaamde ongestructureerde data, zoals e-mail⁶ en webpagina's⁷ en het wordt duidelijk dat er inmiddels een enorme berg data beschikbaar is. Om die berg ook te kunnen 'behappen' zullen data sneller beschikbaar moeten komen. Gartner denkt dat in 2010 meer dan 30 procent van de BI-toepassingen plaats vindt op data die hoogstens 15 minuten oud zijn. (waarschijnlijkheid van 0,7)⁸. Concepten als real-time of active datawarehousing spelen hierop in. Een leverancier als Teradata, die nu beschouwd wordt als high-end leverancier van databases⁹, zou binnenkort weleens mid-end leverancier kunnen worden. Immers, de schaalbaarheid van Teradata is interessant voor organisaties met veel data, veel gebruikers en complexe query's. De huidige ontwikkelingen zorgen ervoor dat een veel groter deel van de markt dezelfde kenmerken heeft.

5. Information democracy.

Met Information democracy wordt bedoeld dat in principe iedereen binnen de organisatie toegang heeft tot alle data. Het concept komt oorspronkelijk uit de koker van Gartner (Howard Dresner), maar is sindsdien door verschillende leveranciers omarmd (Business Objects, Hyperion, Information Builders, etcetera). Terwijl het meestal als keuze wordt gepositioneerd blijkt information democracy in praktijk vaak een gedwongen ontwikkeling. Organisaties die over steeds meer data beschikken (punt 3) en steeds sneller tot besluiten moeten komen (punt 4), kunnen niet anders dan besluitvorming op een zo laag mogelijk niveau in de organisatie neerleggen. Wie dat niet doet schuift verplicht alle data eerst naar boven (management), waarna ze weer naar alle andere geledingen van de organisatie worden doorgeschoven. Dat kost tijd – veel meer tijd dan beschikbaar is. Information democracy betekent ook dat schaalbaarheid een issue is. Schaalbaarheid heeft hierbij betrekking op de hoeveelheid data, maar ook op het aantal gelijktijdige gebruikers. Leveranciers

van BI-software zijn bezig andere licentie- en prijsmodellen te bedenken, aangezien tegen de huidige prijzen (geënt op een beperkt aantal gebruikers) weinig organisaties zich information democracy kunnen veroorloven.

6. Technologie.

Een van de belangrijkste technologische ontwikkelingen is de overgang van 32-bit architectuur naar 64-bit. In essentie betekent dit dat meer data adresseerbaar zijn en in geheugen gelezen kunnen worden. Dit scheidt de mogelijkheid om af te stappen van fysieke kubussen en in plaats hiervan te werken met virtuele kubussen, die ter plekke in het geheugen worden opgebouwd en ververst. De gebruiker kan nu zelf de benodigde kubus definiëren en naar hartelust aanpassen.

De impact hiervan kan niet overschat worden. Waar kubussen altijd door IT moesten worden klaargezet om vervolgens door de business gebruikt te kunnen worden is het nu de business zelf die de kubussen bouwt die ze nodig heeft. Hiermee verschuift het ontkoppelpunt tussen business en IT en wint de organisatie aan snelheid en slagvaardigheid. Behalve snelheid en aanpassingsvermogen dankzij 64-bit is een tweede voorwaarde dat de BI-software nu eens eindelijk een meer intuïtieve interface krijgt. Wie naar een product als QlikTech kijkt (dat overigens ook het datawarehouse in geheugen wil bouwen) of het eveneens Zweedse Spotfire, ziet dat ook op dat terrein inmiddels verbeteringen zijn geboekt.

7. Search engines.

Steeds meer BI-leveranciers (Business Objects, Cognos, Hyperion, Information Builders, SAS) werken samen met search engines zoals Google, Autonomy of FAST. De bedoeling is om de zoektechnologie te combineren dan wel te integreren in de betreffende BI-tooling. Natuurlijk kun je het als winst zien als de vertrouwde zoekomgeving nu ook van toepassing is (of wordt aangetroffen in) de gebruikte BI-tooling. Een innovatieve oplossing is dit echter niet. Het nadeel van de search engines is dat ze ellenlange lijsten genereren waaraan een gebruiker weinig heeft. Uit onderzoek blijkt dat de meeste gebruikers niet verder kijken dan de eerste pagina zoekresultaten. In plaats van je te concentreren op zoeken zou het veel nuttiger zijn om te focussen op vinden. Dat zou betekenen dat eerder aansluiting gezocht wordt bij kennismanagement. Producten van Collexis, Endeca, Infolution en Philips Media Lab (Aquabrowser) laten zien hoe vinden centraal kan staan.

8. DWH appliances.

Niet alleen aan de voorkant (BI) staat het Inmonplaatje onder druk. Ook aan de achterkant (het datawarehouse, DWH) zijn ontwikkelingen aan te wijzen die het oude DWH concept ter discussie stellen en er een nieuw concept voor in de plaats stellen. Het gaat dan om een categorie van tools die als DWH appliances worden aangeduid. De bekendste leveranciers in deze categorie zijn (in alfabetische volgorde): Bizgres, Datallegro, Kognition en

Netezza. In Nederland en België nog geen bekende namen, maar het betreft wel een ontwikkeling om in de gaten te houden. DWH appliances hebben een aanzienlijk lagere prijs doordat ze zich beperken tot de meest voorkomende functionaliteit. De inzet van goedkope hardware en (vaak) open source databases versterkt dit nog. De beperktere functionaliteit zorgt er ook voor dat het product als minder complex wordt ervaren. Uiteraard zijn de appliances niet voor elke situatie de optimale oplossing. In complexe situaties zul je een oplossing moeten vinden voor de ontbrekende functionaliteit. Ook in deze situatie zal de DWH appliance kostentechnisch interessant moeten zijn.

9. Masterdata en metadata.

Het begrip metadata staat al langer in de belangstelling, maar net zoals elk ander kwaliteitsaspect wordt er veel over gezegd, maar niet veel aan gedaan. Of het begrip masterdata hetzelfde lot is beschoren valt te bezien. Masterdata zijn referentiedata zoals klantdata, productcategorieën, marktindelingen etcetera. Deze data worden in verschillende systemen gebruikt en zouden in principe gelijklopend moeten zijn. In de praktijk is dat vaak niet het geval. Hoewel masterdata management in feite niets met datawarehousing te maken heeft, wordt het wel hier neergelegd. Verschillende leveranciers, waaronder Hyperion, IBM, Kalido en SAP beschikken inmiddels over software voor masterdata management.

10. Service Oriented Architecture.

Leveranciers zijn bezig om hun software langzaam maar zeker op SOA-leest te schoeien. De voordelen van een SOA zijn voor de hand liggend: gedefinieerde services kunnen conform de regels die zijn vastgelegd worden benaderd en gebruikt. Het is dus niet meer nodig om dezelfde functionaliteit meerdere keren te bouwen. Voorbeeld van een SOA is Teradata's Application Platform (TAP), waarmee webservices gedefinieerd kunnen worden die op de

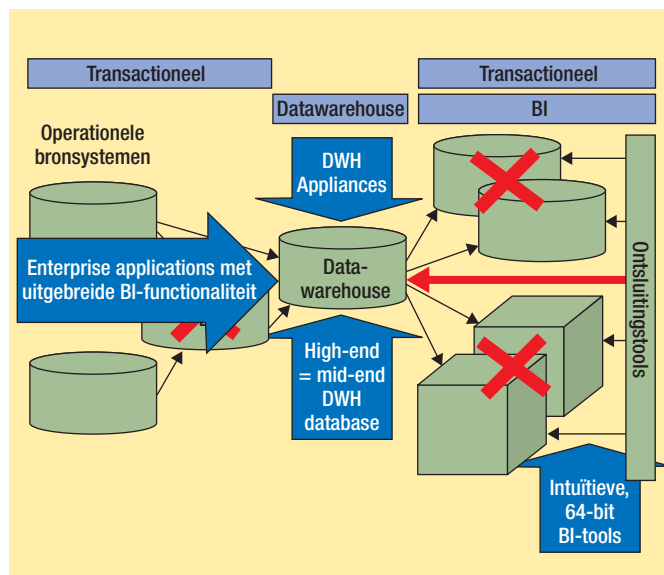
database engine werken. Een ander voorbeeld is Microsoft Office 2007 waarin Excel een webservices-formaat heeft. Tijdens de gebruikersbijeenkomst 2006 kondigde Information Builders de integratie aan van haar BI-platform WebFOCUS en haar Service Oriented Architecture middleware die onder het iWay-label wordt aangeboden. Soortgelijke ontwikkelingen zijn ook bij andere leveranciers waar te nemen.

Conclusie

De belangstelling voor Business Intelligence is nog steeds groeiende. De ontwikkelingen op dit gebied volgen elkaar snel op en versterken elkaar. De exponentiële datagroei en kortere beslissingstijd dwingen organisaties tot information democracy. Information democracy (iedereen in de organisatie heeft toegang tot alle data) dwingt BI-leveranciers tot ander prijs- en licentie-modellen. Door koppeling met search engines moet BI ook voor iedereen vindbaar blijven.

Doordat organisaties het aantal leveranciers waarmee ze zaken doen willen beperken, blijven overnames plaatsvinden. Technologische ontwikkelingen zoals de overgang naar 64-bit technologie brengen tegelijkertijd nieuwe BI-vendors voort die het bestaande Inmon-model onder druk zetten. In datawarehousing zijn het de appliances (alleen de meest gebruikte functionaliteit, tegen een lagere prijs) die voor druk zorgen. De immense hoeveelheden data, groeiende aantallen gebruikers en de vraag om snellere informatie zorgen ervoor dat een high-end database als Teradata binnenkort wel eens mid-end zou kunnen zijn.

Achter al deze ontwikkelingen zit de structurele beweging dat de focus van BI en datawarehousing langzamerhand verschuift van opslag van data (technisch) naar het gebruik van de informatie (business). BI volwassen? Nog lang niet. Maar de contouren van het nieuwe BI en datawarehousing worden langzamerhand duidelijk.



Afbeelding 3: Contouren van de nieuwe BI en datawarehousing.

Noten

1. 'Worldwide Data Warehousing Tools 2005 Vendor Shares', excerpt, Dan Vessel (IDC, 2006).
2. Database Magazine 5, september 2006.
3. 'Hype Cycle for Business Intelligence and Corporate Performance Management, 2006', Andreas Bitterer et.al. (Gartner, 14 July 2006).
4. 'Honger of trek?', Guus Pijpers, TIEM 15, september 2006.
5. Gegevens van de Winter corporation (www.wintercorp.com).
6. Jaarlijks worden 9000 miljard e-mails verstuurd ('Honger of trek?', Guus Pijpers, TIEM 15, september 2006).
7. Het internet groeit met 10 miljoen pagina's per dag ('Honger of trek?', Guus Pijpers, TIEM 15, september 2006).
8. 'Survey shows BI Users want fresher Data', Bill Gassman, Kurt Schlegel, Mark A. Beyer (Gartner, 11 september 2006).
9. Gartner ziet Teradata als marktleider voor datawarehouse database management-systemen (Gartner Magic Quadrant september 2006) en in datawarehouse DBMS servers (Gartner Magic Quadrant augustus 2006).

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).