



Teradata komt met nieuwe Platform Family

Datawarehousing-markt rijp voor consolidatie?

Frans Godden

Eind juli nam Microsoft datawarehousing-leverancier DATAlegro over, volgens analisten mogelijk de start van een consolidatiegolf in dit marktsegment. Wie zich daar geen zorgen over maakt is marktleider Teradata, al jaren op nummer één in het Magic Quadrant van Gartner en sedert zijn partnership met analytics-leverancier SAS sterker gewapend dan ooit in de concurrentiestrijd.

De acquisitie van DATAlegro door Microsoft zorgde toch voor een kleine schokgolf in de datawarehousing-markt, vooral omdat Microsoft hiermee niet alleen een grote lacune in zijn gamma opvulde maar ook omdat het volgens analisten op die manier Oracle de pas afsneed in het hogere database-segment. Het lag dan ook voor de hand dat Teradata hierover aan de tand gevoeld zou worden tijdens ons bezoek aan hun R&D-centrum in Rancho Bernardo in Zuid-Californië. Op onze vragen kwam een direct en eerlijk antwoord van Chris Twogood, marketingdirecteur bij Teradata: DATAlegro heeft een interessante technologie, maar hun probleem was dat ze te weinig visibiliteit hadden in de markt, en dat bleef voor hen de grootste uitdaging. "Ze blijven wat ter plaatse trappelen in de niche van decision support, maar met Microsoft krijgen ze nu misschien wel de kans om door te stoten. De vraag is alleen hoe Microsoft hen gaat gebruiken. DATAlegro is sterk direct sales georiënteerd terwijl Microsoft eerder het 'one to many' concept aankleeft. Het kan dan ook best een hele tijd duren vooraleer we praktische resultaten zien van die overname, algemeen wordt verwacht dat Microsoft de technologie van DATAlegro eerst en vooral in zijn SQL Server database zal willen integreren", aldus Twogood.

Landschap quasi-ongewijzigd

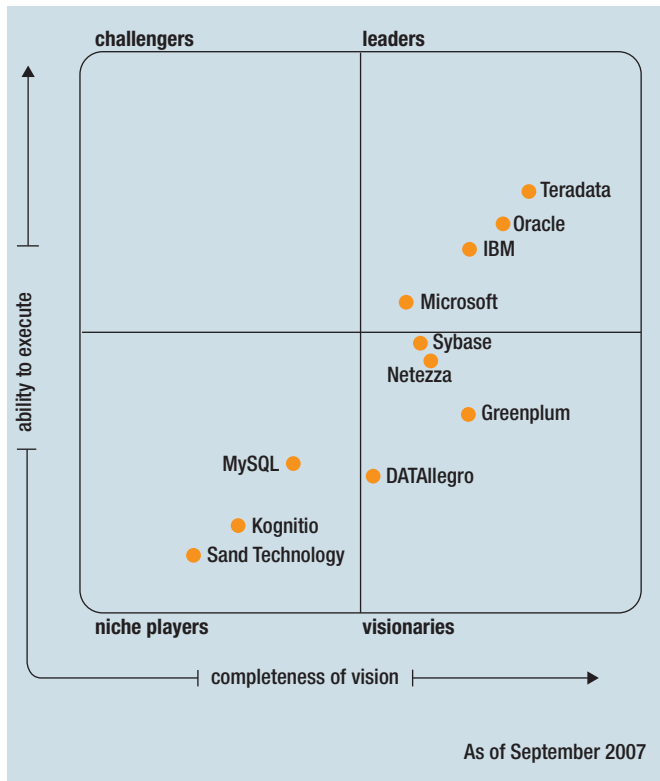
Twogood maakt zich dan ook geen zorgen om dat samengaan, voor hem blijven vooral Oracle en IBM de twee grootste concurrenten in het Enterprise Data Warehousing segment, met HP Neoview en Netezza een trapje lager. Over die laatste twee heeft hij trouwens een uitgesproken mening die volgens hem gebaseerd is op klantencontacten. "Met Neoview is HP veel te snel op de markt gekomen als een product dat alles moest doen wat wij kunnen maar dan goedkoper, maar het is duidelijk dat het product nog niet klaar is. En het probleem met Netezza is dat hun oplossing eigenlijk niet schaalbaar is, op een bepaald

moment loop je tegen het plafond aan en dan moet je, zo vertellen de klanten, alles overboord gooien en overschakelen naar grotere spelers als Teradata, Oracle of IBM".

Dat probleem heb je sedert april van dit jaar nooit meer met Teradata, zo stelt Twogood met klem. Toen introduceerde het bedrijf immers zijn 'Platform Family' met drie grote productgroepen die elk specifiek op bepaalde analytische noden afgestemd zijn. Allemaal maken ze gebruik van de Teradata 12 Database. Onderaan het platform staat de Teradata 550 die vooral bedoeld is voor departementale toepassingen en test- en ontwikkelingsomgevingen tot 6 TB, met een prijskaartje vanaf 67.000 dollar per TB.

Door het gebruik van Teradata 12 kan men probleemloos meegroeien en migreren

Een trapje hoger zit de 2500 die volgens Teradata geschikt is voor entry-level datawarehousing en analytical sandboxes, tot 140 TB aankan en 119.000 dollar per TB kost. Het systeem is geoptimaliseerd voor hoge DSS-workloads en diepgaande analytische toepassingen, maar het mist wel de high availability en het beheer van mixed workloads om een echt Active Enterprise Data Warehouse te kunnen zijn waarmee honderden gebruikers en meerdere toepassingen ondersteund kunnen worden. Daartoe kan Teradata dan zijn 5500/5550 inzetten die niet terugschrikt voor 4 Petabytes en vanaf 200.000 dollar per TB beschikbaar is.



Afbeelding 1: Magic Quadrant for Data Warehouse DBMS.

Bron: Gartner 2007.

Twogood: "Het mooie aan onze Platform Family is dat je door het gebruik van Teradata 12 probleemloos kan meegroeien en migreren – dezelfde database, dezelfde query's, tools, training, BI-processen, noem maar op, zonder dat je ETL, datamodellen of de onderliggende structuren hoeft te wijzigen". Al geeft hij even later wel toe dat je de hardware niet zomaar lineair kan uitbreiden – als je de 550 ontgroeit, moet je een 2500 kopen. Idem voor de volgende groeifase naar de 5500. Maar je toepassingen blijven onaangeroerd verder draaien binnen Teradata 12, en dat is natuurlijk een sterk punt.

Met SAS in zee

De kern van elke degelijke datawarehousing-toepassing is nog altijd de datamining-component. Met datamining kan je immers in je data patronen ontdekken en interpreteren die je in staat stellen problemen op te lossen, en in die optiek bepaalt datamining dan ook de intelligentie van je analytische toepassingen. "Analytische toepassingen zijn maar zo goed als de regels en de kennis die ze aangeleverd krijgen", zegt Arlene Zaima, advanced analytics program manager bij Teradata. "Zorgt je datamining voor het juiste inzicht en diepgaande kennis, dan kan je het gedrag van je klant beter doorgronden, een betere relatie uitbouwen en je klant ook behouden. En dat wil iedereen wel. Helaas is datamining geen eenvoudige klus die je zomaar in een tel kan afronden, het duurt vaak weken of zelfs maanden om analytische modellen te ontwikkelen, want het gaat hier om grote en meestal erg complexe datavolumes die niet zelden de

hardware tot aan zijn grenzen belasten. En dat is een van de redenen waarom wij eind vorig jaar ons partnership met SAS hebben aangekondigd".

En waarom SAS en niet Business Objects of Cognos of Hyperion? "Heel eenvoudig", zegt Zaima, "SAS is de marktleider op het vlak van analytics, en de meeste van onze klanten maken gebruik van SAS. De voorbije jaren hebben we echter vastgesteld dat wij heel wat kunnen doen om dat gebruik te optimaliseren, want SAS haalt data uit ons datawarehouse en integreert alles dan in zijn data sets voor verdere verwerking. En precies dat integratieproces blijkt erg moeizaam te verlopen, een echte bottleneck. Daarom hebben we gezegd: waarom doen we die verwerking al niet binnen de database en creëren we optimale analytische data die SAS dan alleen uit onze database hoeft te halen – geen tonnen gegevens met alle varianten maar kant en klare data waarmee ze onmiddellijk hun modellen kunnen bouwen".

Uniek partnership

En dus kondigden Teradata en SAS op de Partner 2007 User Conference van Teradata in oktober vorig jaar een strategisch partnership aan om die productintegratie te kunnen waarmaken. "Wij werkten voordien ook al wel samen maar dan ging het eerder om een los technologie-partnership", legt Arlene Zaima uit. "Soms waren we concurrenten, soms waren we complementair. Maar nu hebben we een unieke relatie uitgebouwd die de steun geniet van de CEO's van beide bedrijven die er ook de nodige middelen voor hebben vrijgemaakt en een roadmap hebben uitgetekend. En het is geen louter samenwerking met woorden, integendeel: we hebben gezamenlijke ontwikkelings-teams opgezet – enkele Teradata-specialisten wonen en werken nu in de hoofdzetel van SAS in Cary, North Carolina, en heel wat andere vliegen heen en weer tussen Teradata en SAS.

Een multi-temperature warehouse gaat prioriteiten toekennen op basis van business rules

Het partnership gaat echt heel ver – zo hebben de SAS-mensen rechtstreekse toegang tot onze systemen en wordt er zeer nauw samengewerkt op tal van domeinen zoals databases, Business Intelligence, data-integratie. Onze Master Data Management teams werken samen met het DataFlux-team van SAS, net zoals ons beider datamining teams. De bedoeling is echt tot in-database analytics te komen die IT en analisten in staat stelt sneller en accurater in te spelen op de immer wijzigende noden van de business".

De resultaten hebben alvast niet op zich laten wachten. Vijf maanden na de bekendmaking van het partnership kondigde

SAS op zijn eigen Global Forum User Conference dit voorjaar al de SAS Scoring Accelerator for Teradata aan die het klanten mogelijk moet maken scoringmodellen die in SAS Enterprise Miner gecreëerd zijn automatisch te vertalen naar Teradata-specifieke functies, die rechtstreeks in een Teradata-omgeving kunnen worden verwerkt. Arlene Zaima: "Benchmarks hebben al aangetoond dat op die manier records 45 keer sneller kunnen worden verwerkt dan met traditionele methodes. Je moet ook niet langer de SAS scoringcode manueel gaan vertalen naar SQL, alles gebeurt automatisch. De Scoring Accelerator sluit overigens perfect aan bij onze Teradata Warehouse Miner 5.2 die we onlangs geïntroduceerd hebben en waarmee we datamining-modellen en -methoden als UDF's, user-defined functions, kunnen integreren in Teradata. Die UDF's werken dan als een index waaruit zakelijke gebruikers 'voorverpakte' functies kunnen selecteren die gelinked zijn aan gedetailleerde data, zonder nog tijd te verliezen met het manueel coderen van die functies of het heen en weer 'sassen' van data tussen systemen".

Consumenten krijgen rechtstreeks toegang tot datawarehouses via self-service toepassingen

In de roadmap is voorzien dat in de komende jaren eventueel ook SAS Modeling en SAS Scoring mee opgenomen zouden worden in het Enterprise Data Warehouse van Teradata om de snelheid nog te verhogen. "Want dat is precies wat onze klanten vragen: meer performance, en SAS gemakkelijker kunnen gebruiken binnen een Teradata-omgeving, zonder al te veel dataverkeer en zonder SQL-vertalingen, die willen we verbergen zodat een gebruiker alleen in SAS hoeft te werken terwijl alle conversie onderliggend gebeurt", stelt Zaima.

Concept verder doortrekken

Technisch was dit volgens Zaima allemaal niet zo moeilijk, het vergde alleen veel tijd en middelen – "maar het is zoals met elk R&D-programma; als je er de nodige middelen voor kan vrijmaken, dan komen de resultaten vanzelf. Overigens willen we dit concept van analytische data sets die rechtstreeks in het datawarehouse zitten, ook doortrekken naar andere leveranciers waarmee onze klanten zouden werken. Zoals een SPSS of zelfs hun eigen tools, waarbij wij ons concentreren op de data-voorbereiding en de gebruikers er dan hun geoptimaliseerde data kunnen uithalen".

De reacties op het partnership zijn volgens Zaima in elk geval erg positief. Vooral het gezamenlijke Center of Excellence dat beide bedrijven hebben opgericht, kan op zeer veel belangstelling rekenen. Klanten kunnen hier terugvallen op een

speciale ploeg van solution architects en technical consultants om hun IT-infrastructuur te optimaliseren. Zaima citeert een grote verzekeringsmaatschappij (van wie ze de naam nog niet mag noemen) die met behulp van het Center of Excellence de verwerking van gegevens met een SAS Enterprise BI-tool en Teradata van meer dan 40 minuten naar minder dan een minuut heeft teruggebracht.

En wat met andere leveranciers van Business Intelligence, blijven die in de kou staan? "Nee, toch niet", stelt Zaima met klem, "wij werken met vele BI-leveranciers samen om ook hun technologie te optimaliseren. Zo bijvoorbeeld heel wat met Microsoft dat nu een services cube heeft aangekondigd die gegevens in een Teradata database zou opslaan in een ROLAP-model, Relational OLAP dus. Maar ons partnership met SAS blijft wel uniek".

Van active naar multi-temperature

Teradata propageert intussen al jaren het concept van een Active Data Warehouse waarmee je een einde kan maken aan de vertragingen in de datastroom die operationele beslissingen blijven afremmen. "Wat wij doen is de tijd die verstrikt tussen de registratie van een gebeurtenis en de informatie die daaruit voortspruit verkorten, zodat je sneller kan reageren op basis van die gegevens – of met andere woorden: Active Enterprise Intelligence op basis van real-time informatie", zegt Scott Gnau, chief development officer bij Teradata.

Volgens chief technical officer Todd Walter werkt Teradata intussen volop verder aan een ander concept dat vorig jaar voor het eerst aangekondigd werd, multi-temperature warehousing. Daarbij is 'temperatuur' een metafoor voor de frequentie waarmee data geraadpleegd worden – hoe vaker dat gebeurt, hoe hoger de temperatuur van die data ligt. "Het concept is eenvoudig", legt Walter uit. "Een multi-temperature warehouse gaat prioriteiten toekennen aan het gebruik van de beschikbare systemen op basis van business rules en tegelijkertijd het gebruik van de opslagcapaciteit optimaliseren. De temperatuur wordt daarbij bepaald door vier factoren: de frequentie van de raadpleging, van de updates, het onderhoud, en de data-historiek. Teradata's workload management maakt het mogelijk om niet alleen actieve – hot – data maar ook inactieve – koude – data mee in het datawarehouse te beheren. Dat gebeurt met behulp van onze Virtual Storage Services, die in staat zijn snellere en tragere sectoren op schijven te herkennen en data dan aan bepaalde sectoren toe te wijzen".

Todd Walter twijfelt er niet aan dat een multi-temperature datawarehouse de basis zal vormen voor de volgende generatie van analytische toepassingen, alleen al omdat de datavolumes zo enorm zullen toenemen en steeds meer consumenten rechtstreeks toegang zullen krijgen tot datawarehouses via allerlei self-service toepassingen.

Frans Godden is freelance journalist.