



SQL Server 2005: moeten BI-leveranciers zich zorgen maken?

Belangrijke tussenstop op weg naar BI-platform

Paul van der Linden

Na aanzienlijke vertraging heeft Microsoft op 7 november de lang verwachte opvolger van SQL Server 2000 gelanceerd. MS SQL Server 2005 (eerder bekend onder de codenaam Yukon) biedt met name verbeteringen in Business Intelligence en meer integratie in software-ontwikkeling. Ook de database zelf heeft een aantal belangrijke verbeteringen ondergaan.

Een van de aspecten waar Microsoft aan heeft gewerkt is de verdere uitbouw van BI-functionaliteit. Nu staat Microsoft niet te boek als een BI-vendor. Anderzijds is het wel een partij die in het verleden al heeft bewezen om markten beslissend te kunnen veranderen. Hoeveel BI zit er in SQL Server 2005? Hoe toegewijd is Microsoft aan BI? En: hoe bevreesd moeten BI-vendors zijn voor de leverancier uit Redmond?

Achtergrond

Microsoft behoeft uiteraard geen introductie. Er zullen weinig personen zijn die niet te maken hebben met (of tenminste op de hoogte zijn van) Microsoft Office-producten of het Windows besturingssysteem. Dit zijn zeker niet de enige activiteiten die de software-reus verricht (denk bijvoorbeeld aan de Xbox, Axapta, Outlook of Sharepoint), maar het zijn wel de activiteiten waarop Microsoft drijft en die het merendeel van de inkomsten verzorgen. Binnen de Office-producten zijn het met name de spreadsheet Excel en de tekstverwerker Word die het meest gebruikt worden. Business Intelligence is nooit het belangrijkste speerpunt voor Microsoft geweest. De SQL Server database (in het verleden ontstaan uit een gezamenlijke ontwikkeling met Sybase) kent weliswaar de multidimensionale database Analysis Services (AS) – voorheen bekend als OLAP Services – en heeft een eigen front-end tool (Data Analyzer), maar Excel is eigenlijk Microsofts meest populaire BI-tool. Uit het Nationaal Data Warehouse Onderzoek dat Array eerder dit jaar in samenwerking met Atos Origin uitvoerde bleek nog eens dat Excel de meest gebruikte rapportage-tool is en ook behoort tot de top 5 van meest gebruikte OLAP-tools. Ontsluitings-tools voor Analysis Services worden geleverd door leveranciers zoals Panorama, ProClarity en Temtec. Dat Microsoft over voldoende diepe zakken beschikt om volwaardige software op te hoesten zal niemand betwisten. Waarbij de algemene opinie is dat Microsoft drie pogingen nodig heeft om

succesvol te zijn. Aangezien dit scenario op BI-gebied (nog) niet aan de orde is lijkt het erop dat Microsoft BI nog onvoldoende interessant vindt. Of is hier iets anders aan de hand?

MS SQL Server

Microsoft positioneert SQL Server 2005 als datamanagement- en analyse-software gericht op volledige en kostenefficiënte analyse en rapportage. Het is geen toeval dat de introductie van SQL Server 2005 samenviel met die van VisualStudio 2005 en BizTalk Server 2006. De samenwerking tussen deze drie producten wordt steeds groter. Microsoft lijkt dan ook bezig te zijn om een heus BI-platform in te richten.

Aan de SQL Server database zijn in het verleden al een OLAP-oplossing (OLAP Services) en een ETL-product (Data Transformation Services – DTS) toegevoegd. OLAP Services werd reeds eerder getransformeerd naar het huidige Analysis Services. Het bevat naast OLAP-functionaliteit ook mogelijkheden om aan datamining te doen.

Hierna is Reporting Services toegevoegd – een rapportage-tool die met name bedoeld lijkt voor ontwikkelaars en minder voor eindgebruikers. In de stap naar SQL Server 2005 is DTS ingewisseld voor Integration Services. De ontwikkeling die Microsoft hiermee inzette is die van een 'aangeklede' database. Behalve de reguliere database-functionaliteit worden ook analyse, rapportage en ETL-functionaliteit geboden. Inmiddels bieden de databases van Oracle en IBM ook additionele functionaliteit.

Verbeteringen en vernieuwingen

Wat zit er in de doos die SQL Server 2005 heet? De verbeteringen en nieuwe elementen in SQL Server 2005 betreffen:

- End-to-end integratie van het BI-platform;
- Integration Services (voorheen: Data Transformation Services – DTS);

- Analysis Services;
- Reporting Services;
- Integratie met de MS Office-producten.

Hierna worden enkele van deze aspecten nader toegelicht.

Ontwerp en data-integratie (Integration Services).

Nieuw is de Business Intelligence Development Studio. Van hieruit kunnen op visuele wijze mappings worden gemaakt waarmee data vanuit verschillende bronnen naar het datawarehouse kunnen worden gehaald. BI Development Studio is gebaseerd op Visual Studio en maakt gebruik van Analysis Services, Reporting Services en de database engine. Met behulp van BI Development Studio worden SQL Server Integration Services (SSIS) packages gemaakt.

SQL Server 2005 heeft Data Integration Services ingeleverd voor SSIS. Hiermee komen complexe ETL-bewerkingen en het verwerken van grote hoeveelheden data binnen bereik.

Analyse (Analysis Services).

Microsoft Data Mining is uitgebreid met een aantal nieuwe algoritmes, waaronder Association Rules, tijdreeksen, Regression Trees, Sequence Clustering, neurale netwerken en Naïve Bayes. Het onderscheid tussen relationele en multidimensionale opslag is meer naar de achtergrond verschoven. De Proactive Cache feature biedt de mogelijkheid om aan te geven wanneer een kubus moet worden ververs in geval van gewijzigde brondata. Hiermee kan zelf de afweging tussen performance en actualiteit van data worden gemaakt. Analysis Services cubes hebben daarnaast additionele functionaliteit gekregen waaronder het key performance indicator (KPI) framework, MDX-scripts en andere embedded business analytics.

Analysis Services biedt een zogenaamd Unified Dimensional Model (UDM). Dit is een verzameling van kubussen en dimensies die dienst doen als centrale metadata repository voor business-entiteiten, business-logica, berekeningen en metrics. Vanuit deze repository worden alle rapporten, spreadsheets, OLAP browsers, KPI's en analytische applicaties gegenereerd. Analysis Services

bestaat uit twee complementaire stukken software: data mining en OLAP. Beide vormen de basis voor de analytische applicaties. Door gebruik te maken van Data Source View (ook een nieuwe feature) kan UDM gemapt worden op de verschillende bron-systemen. Hierdoor kan een geïntegreerd beeld van de business worden verkregen. Het Unified Dimensional Model kan worden gebruikt om de corporate data te onderzoeken.

Analysis Services (AS) ondersteunt XML for Analysis (XML/A) en biedt meer support voor MDX. MDX-scripts zijn de nieuwe manier om calculated members, named sets en cell calculations te definiëren. De vroegere AS metadata repository is in deze versie vervangen door XML-files die worden opgeslagen en beheerd door AS. XML is nu de data definition language (DDL) voor Analysis Services. XML/A is de standaardwijze waarop AS

Het Unified Dimensional Model kan worden gebruikt om de corporate data te onderzoeken

communiceert (web services).

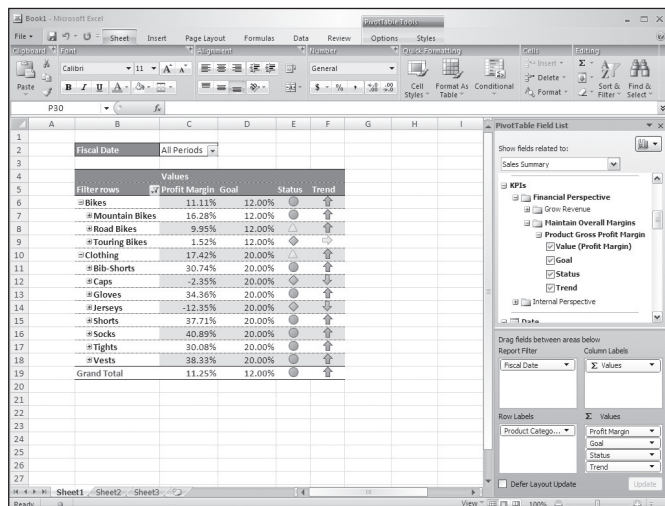
In tegenstelling tot de vorige versie (AS 2000) vinden nu alle berekeningen centraal op de server plaats. De client-side cache bestaat niet meer, waardoor een zero footprint-oplossing wordt bereikt. Het beoogde effect is dat het uitvoeren van complexe berekeningen hierdoor sneller gaat. De andere kant ervan is dat het uitvoeren van eenvoudige query's wellicht trager gaat (dit moet nu immers via de server lopen).

Rapportage (Reporting Services).

Reporting Services is de component die de creatie en het verspreiden van rapporten verzorgt. Rapporten kunnen worden gepersonaliseerd en op verschillende wijzen worden geformatteerd. Nieuw in SQL Server 2005 is Report Builder; dit verschaft eindgebruikers de mogelijkheid om ad hoc rapporten te maken.

Component	SQL Server 2000	SQL Server 2005
Extract, transformation, and load	Data Transformation Services (DTS)	SQL Server 2005 Integration Services
Relational datawarehouse	SQL Server 2000 relational database	SQL Server 2005 relational database
Multidimensional database	SQL Server 2000 Analysis Services	SQL Server 2005 Analysis Services
Data mining	SQL Server 2000 Analysis Services	SQL Server 2005 Analysis Services
Managed reporting	SQL Server 2000 Reporting Services	SQL Server 2005 Reporting Services
Ad hoc reporting	Not applicable	SQL Server 2005 Reporting Services
Ad hoc query and analysis	Microsoft Office products (Excel, Office Web Components, Data Analyzer, SharePoint Portal Server)	Microsoft Office products (Excel, Office Web Components, Data Analyzer, SharePoint Portal Server)
Database development tools	SQL Server 2000 Enterprise Manager, Analysis Manager, Query Analyzer, various other tools	SQL Server 2005 Business Intelligence Development Studio (New)
Database management tools	Enterprise Manager, Analysis Manager	SQL Server Management Studio (New)

Afbeelding 1: Van SQL Server 2000 naar 2005.



Afbeelding 2: Informatie-integratie met next generation SharePoint.

Dat was ook nodig omdat Reporting Services moeilijk in het gebruik was en gezien de inspanning weinig opleverde. Report Builder maakt gebruik van het Reporting Services platform. Microsoft heeft de opzet en werkwijze van Report Builder zoveel mogelijk conform de MS Office-productlijn vormgegeven. Gestart wordt met een rapport lay-out (template) waar met behulp van *drag and drop* elementen kunnen worden toegevoegd.

Manage (database engine).

SQL Server Management Studio is de software waarmee het hele proces kan worden aangestuurd. Waar BI Development Studio er is voor het creëren van mappings etcetera, richt Management Studio zich op het beheer en onderhoud ervan. Beide producten moeten gezamenlijk zorgen voor een complete ontwikkel-omgeving om BI-applicaties te maken.

Aan Transact-SQL zijn verschillende nieuwe datatypes en analytische functies toegevoegd. Voor datawarehouse-gebruik zijn de datatypes varchar(max), nvarchar(max) en varbinary(max) interessant. Elk kan tot 2 Gigabyte aan data opslaan en hiermee vormen ze een alternatief voor text-, ntext- en image-datatypen. Met name waar veel beschrijvende tekst of metadata moeten worden opgeslagen zijn dit nuttige mogelijkheden.

Tot de nieuwe analytische functies behoren: row_number, rank, dense_rank en ntile. In een geordende resultaatset deelt row_number aan elke rij een hoger nummer uit. Rank houdt rekening met identieke waarden. Als rij 1 en 2 een identieke waarde hebben zal het row_number 1,2 en vervolgens 3 zijn. Rank levert hier 1,1 en 3 op (rank 2 wordt overgeslagen). Dense_rank levert in deze situatie 1,1 en 2 op. Dit is nuttige functionaliteit, met name in een data staging area. Ntile verdeelt een geordende set in het gespecificeerde aantal (n) groepen van bij benadering gelijke grootte.

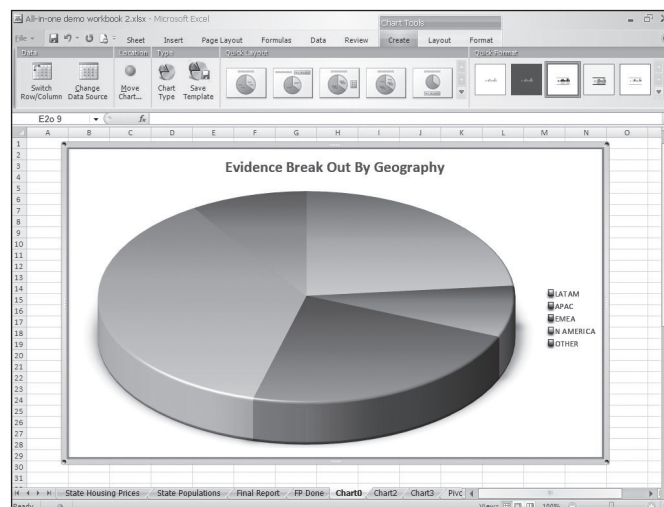
SQL Server in verschillende smaken

Microsoft biedt SQL Server 2005 aan in een aantal mogelijke varianten: Express, Standaard, Workgroup en Enterprise. Express

is bedoeld om kennis te maken met de geboden functionaliteit en is vrij verkrijgbaar. Er zit een aantal beperkingen bij, maar de meeste functionaliteit is aanwezig om uitgeteerd te worden. Partitionering, parallelle indexen en index views zijn in deze variant niet beschikbaar. De grootte van de database is beperkt tot 4 GB. Voor de overige varianten geldt geen beperking op database-grootte. Hierbij worden ook situaties met meerdere processoren ondersteund (onbeperkt bij de Enterprise-variant). Overigens is de Enterprise-variant ook de enige waarbij partitionering, parallelle indexen en index views ondersteund worden. Prijzen zijn gebaseerd op het aantal processoren of op de combinatie van server en gebruikers/apparaten (CAL) die gebruik maken van SQL Server. Dat betekent dat de prijs voor SQL Server Enterprise neerkomt op 25.000,- dollar of 8.487,- dollar per server en 162,- dollar per CAL (indicatieve prijzen).

Microsoft BI apps

Microsoft heeft recentelijk in november de Microsoft Office Business Scorecard Manager 2005 gelanceerd. Dit is een server-gebaseerde Business Performance Management (BPM) applicatie die de Business Scorecards Accelerator opvolgt die in 2004 werd geïntroduceerd. Hiermee kunnen key performance indicatoren (KPI's) en business scorecards worden gedefinieerd en toegepast. Dit alles gebeurt vanuit de vertrouwde MS Office-omgeving. In de volgende versie van MS Office (codenaam: Office '12') zal Excel worden omgeturnd tot een echte BI-tool die beschikt over een persistente connectie met databronnen. Dat zou betekenen dat zaken als consistentie door de database kunnen worden afgehandeld waardoor een groot nadeel van het huidige Excel-gebruik wordt opgelost. Ook is er sprake van serverside Excel wat de naam Excel Services krijgt en zorgt voor het managen, delen en beveiligen van spreadsheets. Het is daarnaast de bedoeling dat Excel '12' volledige ondersteuning van Analysis Services 2005 biedt. Sorteren, filteren, datavisualisatie alsmede PivotTables en PivotCharts worden verder uitgebreid. Sharepoint vormt hierbij het portal voor Reporting Services en de Office-toepassingen.



Afbeelding 3: Microsoft Office Excel '12' – PivotCharts.

Positionering

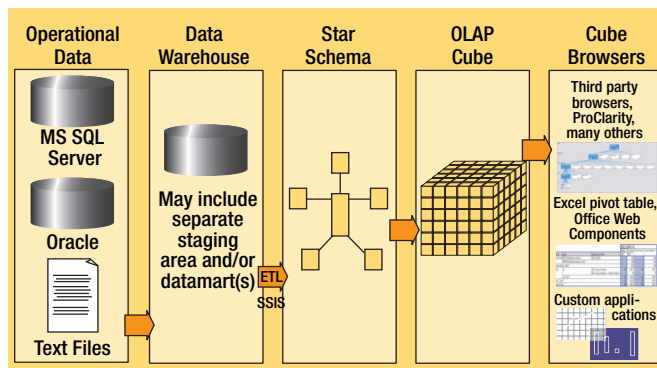
Microsoft is geen pure play BI-vendor. Bekeken vanuit een BI-invalshoek kan Microsoft nog het best geschaard worden in een zelfde categorie als Oracle en IBM. Alle drie zijn database-leveranciers die op die wijze ook BI leveren. Waarbij Microsoft nog het meest vanuit de database vertrekt. In SQL Server is langzamerhand steeds meer BI-functionaliteit opgenomen. Een beweging die niet onopgemerkt is gebleven en navolging heeft gekregen bij andere database-leveranciers. Zo bekeken lijkt Microsoft een tamelijk grijze speler in de BI-markt en zeker geen grote bedreiging voor spelers zoals Business Objects, Cognos, Hyperion, Information Builders of SAS.

Maar wie verder kijkt kan alleen maar tot de conclusie komen dat de stealth aanpak van Microsoft in de BI-markt niet zonder gevolgen is gebleven en dat er een visie achter zit. Het opnemen van het ETL-product Data Transformation Services (nu Integration Services) bij de database heeft de prijsstelling van dit soort (dure) producten aanzienlijk onder druk gezet. En hoewel het niet de meest uitgebreide functionaliteit biedt is het voor veel organisaties zeker voldoende. Geconcludeerd moet worden dat de voormalige ETL-kampioenen Informatica en Ascential (nu IBM) zich nu bezighouden met data-integratie.

Toevoegen van OLAP Services (nu Analysis Services) aan SQL Server was een verdere stap om Business Intelligence tot een gewone activiteit terug te brengen en meer mensen in contact te brengen met OLAP (analyse) en datamining. Ook hier heeft Microsoft een centrale rol gespeeld en wel door het bijeenbrengen van de relationele en multidimensionele database waarbij het onderscheid er tussen steeds diffuser en onbelangrijker zal worden (black box). In het geval van de rapportage-server was Cognos met ReportNet eerder en van meer belang. Reporting Services heeft hier nog iets in te halen, maar Microsoft is wel eerder vanuit een achterstand teruggekomen. Waarbij veelal geldt dat ze uiteindelijk misschien niet het beste, maar wel het meest succesvolle product in handen hebben.

Visie

Een groot deel van de data waarop BI wordt toegepast is ongestructureerd. Dit zijn data die niet goed passen in de rijen en kolommen van een relationele database. Denk dan bijvoorbeeld aan de data aanwezig in e-mails, presentaties of tekstverwerkers. En het zijn juist deze data die grotendeels in Microsoft-applicaties als Outlook, PowerPoint of Word ontstaan. Microsofts visie is om deze data niet te verplaatsen naar aparte BI-tools, maar ter plekke BI erop toe te passen. Enerzijds verhoogt dit de productiviteit van de information worker; anderzijds wordt hiermee de noodzaak voor aparte BI-tools aanzienlijk verminderd. Die zou je dan alleen nog voor hele complexe zaken nodig hebben. De vertrouwde van de MS Office-producten wordt hiermee aangewend om ook Business Intelligence binnen handbereik van een ieder te brengen. Zie hier de visie en strategie van Microsoft als BI-leverancier. Als het werkt kan het een uitstekende oplossing zijn. En heeft Microsoft haar BI-concurrenten verslagen door de spelregels te



Afbeelding 4: Analysis Services Architecture.

veranderen. Invulling geven aan deze visie is echter niet zo makkelijk. Ja, Excel is wellicht de meest gebruikte BI-tool. Maar het staat tegelijkertijd bekend als onuitroeibaar virus. En dan gaat het om de kosten die ongebreideld en niet-gereguleerd gebruik ervan een organisatie opleveren. Aan de beoogde invulling wordt dus nog gewerkt en SQL Server 2005 is weer een belangrijke stap in deze richting. Het is nog niet de hele invulling van de visie – en tot die tijd is er nog ruimte voor BI-leveranciers die aanvullende functionaliteit leveren.

Conclusie

Microsoft wordt niet direct gezien als een BI-leverancier – en dat is ook niet de manier waarop het bedrijf uit Redmond zich positioneert. Maar dat betekent niet dat het op dit gebied niets te bieden heeft of zich niets aan BI gelegen laat. De nieuwste versie van haar database SQL Server 2005, de langverwachte opvolger van SQL Server 2000, biedt dan ook interessante uitbreidingen en aanvullingen op Business Intelligence gebied. SQL Server Management Studio, Business Intelligence Development Studio, Report Builder, ondersteuning van XML/A en ga zomaar door – we hebben lang moeten wachten, maar het is de moeite waard geweest en het zal zeker weer de standaard zetten voor wat moderne databases aan functionaliteit bieden.

Microsoft smeedt steeds meer software-producten aan elkaar om tot een indrukwekkend BI-platform te komen. SQL Server 2005 is daarin slechts een tussenstop – zij het een belangrijke. In de ogen van Microsoft is BI te duur, te complex en staat het los van de dagelijkse gang van zaken. Haar bedoeling is dan ook om de information worker vanuit de bekende MS Office-omgeving de beschikking te geven over intuïtieve BI-functionaliteit. Naast MS Office en SQL Server spelen ook andere producten zoals Sharepoint, VisualStudio en Business Scorecard Manager daar een belangrijke rol in. De volgende versie van MS Office (Office '12') biedt zelfs een Excel die is omgeturnd tot een officiële BI-tool. Dat lijkt het moment waarop BI-leveranciers zich serieus zorgen moeten gaan maken.

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).