



Analytische MPP-databases worden belangrijker

Aster Data, Greenplum en Vertica

Paul van der Linden

Nu de markt voor datawarehouses gedomineerd wordt door de zogenaamde mega vendors (Microsoft, IBM, SAP en Oracle) kan de indruk ontstaan dat er in die markt dan ook weinig gebeurt. Niets is minder waar. Groeiende hoeveelheden data en gemengde werklasten (mixed workloads) hebben de vraag opgeworpen hoe een acceptabele respons gegarandeerd kan worden.

Gartner gaat ervan uit dat maar liefst 70 procent van alle databases een performanceprobleem heeft. Nu datawarehouses steeds vaker een bedrijfskritisch karakter hebben, de geboden data steeds actueler moeten zijn en ook operationele BI ondersteund moet worden, is performance opeens een cruciaal aspect geworden. Leveranciers zijn dan ook naarstig op zoek naar oplossingen. Voor Aster Data, Greenplum en Vertica geldt dat ze die oplossing alle drie in dezelfde hoek zoeken: die van MPP en MapReduce.

Aster Data, Greenplum en Vertica zijn leveranciers van analytical databases die over het algemeen niet meteen herkenning oproepen. Toch bestaan ze al een tijdje en hebben ze alle drie een plaats weten te bemachtigen in de meest recente versie van Gartner's Magic Quadrant voor Data Warehouse Database Management (januari 2010)¹.

Behalve dat alle drie de bedrijven hun thuisbasis in de V.S. hebben, maken ze ook alle drie gebruik van een *massively parallel processing* (MPP) oplossing, gebruiken ze MapReduce om grote hoeveelheden data te lijf te gaan en gaat het om relatief jonge, niet-beursgenoteerde bedrijven. Dat laatste geldt overigens niet meer voor Greenplum, dat recentelijk (in juli) door EMC is overgenomen.

Aster Data

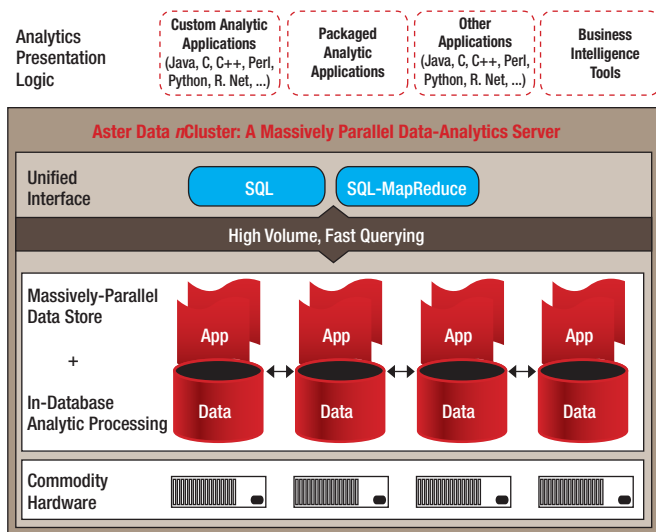
Het doel van Aster Data is om 'advanced analytics' beschikbaar te stellen. Hiermee wordt bedoeld dat enorme hoeveelheden data interactief en ad hoc bevraagd kunnen worden.

De aanpak die Aster Data heeft gekozen is om een MPP-architectuur te combineren met in-database dataverwerking. Het door Google ontwikkelde MapReduce-framework speelt hierbij een belangrijke rol. Eenvoudig gezegd komt MapReduce erop neer dat een vraag opgedeeld wordt in verschillende subvragen die elk door een aparte node worden opgepakt en beantwoord.

De verschillende antwoorden worden dan weer teruggegeven en tot een definitief antwoord geassembleerd. Dat Google het concept heeft neergezet is geen verrassing, aangezien het bedrijf te maken heeft met enorme hoeveelheden data en gebruikers. Verschillende softwareleveranciers waaronder Aster hebben het concept vervolgens geadopteerd en onderdeel gemaakt van hun eigen product.

Steve Woledge, Senior Director of Corporate Marketing bij Aster Data, geeft aan dat de eigen versie van MapReduce SQL-MapReduce heet en native geïmplementeerd is. Dat laatste betekent dat niet met de data wordt geschoven, maar dat de applicatielogica naar de data wordt gebracht. SQL-MapReduce zorgt ervoor dat SQL gebruikt kan worden in combinatie met MapReduce, zonder dat code herschreven hoeft te worden. Hierdoor wordt het mogelijk om zowel SQL als procedurele talen als Java, C++, Python en Perl te gebruiken bij het definiëren van analytische functies en vragen. SQL is relatief eenvoudig en bekend, maar is niet echt krachtig bij het formuleren van analytische functies. Combinatie zorgt dus voor de 'best of both worlds'. Volgens Woledge bieden traditionele datawarehouses slechts een fractie van de beschikbare organisatiedata. Zijn inschatting is dat zo'n 70 tot 80 procent van de data niet in het datawarehouse zit. Een groot deel hiervan bestaat uit zogenaamde ongestructureerde data – door Woledge aangeduid als niet-relationale data. Ook SQL wordt door Woledge als een beperking gezien, reden waarom vanuit analyseoogpunt procedurele talen zeker zo interessant zijn en ondersteund worden.

Het in-database verwerken van data is op zich niet nieuw. Een gevestigde partij als Teradata biedt bijvoorbeeld de mogelijkheid om de functionaliteit van SAS in de database uit te voeren. Het voordeel van het bij elkaar brengen van data en functionaliteit is dat resultaten sneller uitgerekend kunnen worden. Waar voorheen allerlei lagenmodellen uitgingen van het scheiden van data



Afbeelding 1: Aster Data – nCluster overview.

en applicaties (en presentatie), gaan nu steeds meer leveranciers ertoe over om data en applicatie tegen elkaar aan te zetten.

Versies

Aster Data levert een product in drie verschillende versies. Naast de 'kale' softwareversie van nCluster, bestaat ook de mogelijkheid om het in de Cloud te gebruiken en is er een appliance-versie. Hiermee is Aster verrassend compleet. De Cloudversies zijn er voor het Amazonplatform en voor AppNexus. Als appliance wordt nCluster geleverd op hardware van Dell met de data-integratiesoftware van Informatica en MicroStrategy voor BI-toepassingen.

De *sweet spot* voor Aster Data is het marktsegment waar organisaties beschikken over enorme datavolumes gekoppeld aan CPU-intensieve bewerkingen. Daar komt de combinatie van de MPP-architectuur en SQL-MapReduce immers het beste tot zijn recht. Toch meent Woledge dat ook bij andere scenario's Aster Data een interessante optie biedt in termen van prijs/prestatie. CPU-intensief gecombineerd met minder dan 1 TB zou volgens hem zo'n combinatie zijn.

Verschillen

Gevraagd naar de verschillen met de overige MPP analytische databases geeft Woledge aan dat bij Vertica wel wordt geschoven met data en dat een kolomgeoriënteerde database zoals van Vertica meer geschikt is voor eenvoudige query's en rapporten die niet rekenintensief zijn. De implementatie van MapReduce door Greenplum is volgens Woledge niet native, waardoor er ook met data wordt geschoven en het een 'disk heavy boxes'-oplossing is.

Momenteel heeft Aster Data zo'n 100 klanten indien ook de downloads van de website worden meegerekend. Als deze buiten beschouwing worden gelaten gaat het om zo'n 35 *core clients*. Tot de klanten van Aster Data behoren onder andere Telefónica, Barnes & Nobles, MySpace, LinkedIn en Intuit. In Europa is er

een handvol klanten. De prijs die klanten voor nCube moeten betalen is afhankelijk van het datavolume dat bevroegbaar is (queryable). Hiervoor geldt een prijs van een aantal dollar per Terabyte. Er is een sales office in het V.K. Support wordt gedaan vanuit de V.S. en kan 24/7 worden geleverd. Het aantal werknemers staat momenteel op 100, de meesten in de V.S. werkzaam.

Greenplum

In San Mateo (California), op minder dan 10 kilometer van het hoofdkantoor van Aster Data, is Greenplum gevestigd. Greenplum bestaat zeven jaar en heeft meer dan 100 klanten (onder andere NYSE, T-Mobile, Nasdaq, Ricoh en O'Reilly). De database waar Greenplum gebruik van maakt (nu in versie 4) is gebaseerd op PostgreSQL, een open source database. Greenplum positioneert haar product als bijzonder schaalbaar. Het kan hierbij tot honderden Terabytes aan data aan. Net als bij Aster Data is de oplossing ook in een cloudopstelling beschikbaar en heet dan Greenplum Chorus. MapReduce wordt ondersteund en databasecode kan op de storage device worden uitgevoerd, hetgeen een verbeterde performance oplevert. Tot de extensies die Greenplum op de PostgreSQL database heeft toegepast behoren onder andere de mogelijkheid om naast rijgeoriënteerde dataopslag ook kolomgeoriënteerde opslag toe te passen en self healing fault tolerance, die bij een opstelling met meer dan 100 servers de TCO acceptabel houdt. De Greenplum database kan in de 3.3-versie opereren onder diverse smaken van Linux (SUSE, Red Hat, CentOS) en Unix (HP, SUN, IBM). Greenplum biedt voor geïnteresseerden een gratis single node database edition aan, die zelfs kan deelnemen als node in Greenplum's Enterprise Data Cloud. Op een x86-server kan dan geoefend worden met MapReduce processing zonder volumebeperking. Uiteraard bestaat de upgrademogelijkheid naar een volwassen multi-node oplossing. Behalve software biedt Greenplum

CLIENT ACCESS & TOOLS	CLIENT ACCESS ODBC, JDBC, OLEDB, etc.	3rd PART TOOLS BI Tools, ETL Tools Data Mining, etc.	ADMIN TOOLS GP Performance Monitor pgAdmin 3 for GPDB
PRODUCT FEATURES	LOADING & EXT. ACCESS Petabyte-Scale Loading Trickle Micro-Batching Anywhere Data Access	STORAGE & DATA ACCESS Hybrid Storage & Execution (Row- & Column-Oriented) In-Database Compression Multi-Level Partitioning Indexes – Btree, Bitmap, etc.	LANGUAGE SUPPORT Comprehensive SQL Native MapReduce SQL 2003 OLAP Extensions Programmable Analytics
GPDB ADAPTIVE SERVICES	Multi-Level Fault Tolerance	Online System Expansion	Workload Management
CORE MPP ARCHITECTURE	Shared-Nothing MPP Parallel Query Optimizer Polymorphic Data Storage™		Parallel Dataflow Engine gNet™ Software Interconnect MPP Scatter/Gather Streaming™

Afbeelding 2: Greenplum Database.

ook de mogelijkheid om te assisteren bij het uitvoeren van analyses. Hiertoe is een Analytics Lab ingericht waar gezamenlijk de tanden kan worden stukgebeten op complexe opgaven.

EMC

De overname van Greenplum door EMC kan voor de databaseverancier grote gevolgen hebben. EMC, van huis uit leverancier van hardware en software voor dataopslag, krijgt met Greenplum een eigen database in huis. De markt voor Business Intelligence,

MPP versus SMP

SMP (*Symmetric Multi-Processing*) is een computerarchitectuur waarbij de verschillende processoren een geheugen delen. Terwijl elke node (processor) beschikt over zijn eigen schijven, hebben ze door het gemeenschappelijke geheugen toegang tot alle harde schijven en de data die zich daarop bevinden. Door de *tight coupling* maakt SMP gebruik van speciale hardware en software. De achilleshiel van een SMP-opstelling bestaat uit het coördineren van het gebruik van het gedeelde geheugen. Naarmate er meer processors worden toegevoegd, nemen de coördinatiekosten meer dan proportioneel toe. Hiermee bestaat er een praktische beperking op het aantal processoren dat je kunt inzetten. En daarmee ook de hoeveelheid gegevens en de complexe bewerkingen die geschikt zijn voor een SMP-machine. Als reactie op de beperkingen van SMP is MPP ontwikkeld.

Het idee achter MPP (*Massively Parallel Processing*) is om een volledig schaalbare oplossing te creëren op basis van commodity processoren, goedkope schijven en geheugen. Idealiter zou het verdubbelen van het aantal processoren moeten leiden tot het verdubbelen van de capaciteit waarover je kan beschikken. In dat geval is er sprake van volledige schaalbaarheid, ook wel aangeduid als een proportionele toename van de beschikbare capaciteit.

Bij MPP is er sprake van individuele nodes, die niets met elkaar delen. Het gemeenschappelijke geheugen (de bottleneck bij SMP) is hier afwezig. De verschillende nodes delen niets met elkaar en hebben elk hun eigen processor, geheugen en schijven. Het verwerken van een opgave (ophalen van gegevens, beantwoorden van een query) kan verdeeld worden over deze verschillende nodes, maar vereist uiteindelijk toch een connectie tussen de verschillende nodes. Een high-speed network interconnect is nodig om de verschillende nodes aan elkaar te koppelen. Als een bepaalde node data nodig heeft die aanwezig zijn bij andere nodes, zal via de interconnect de betreffende SQL worden gestuurd naar deze nodes. Ter plekke en in parallel kunnen vervolgens de relevante data worden geselecteerd en naar de vragende node worden gestuurd.

Voor beide architecturen zijn voor- en nadelen te noemen. In de praktijk worden SMP-machines met name ingezet in de range tot 1 Terabyte (maximaal 3 TB) en bij eenvoudige en gemiddeld complexe query's. Vanaf 5 Terabyte is MPP de betere keuze.

datawarehousing en analyse is een interessante markt omdat hier enorme hoeveelheden data in omgaan. Deze moeten niet alleen worden opgeslagen en beheerd, maar ook nog eens met een fatsoenlijke snelheid aan managers, analisten en eindgebruikers ter beschikking worden gesteld. Een van de manieren waarop leveranciers hierop inspelen is door hardware en software steeds meer op elkaar af te stemmen. Door Greenplum heeft EMC nu ook de mogelijkheid om de eigen hardware en database zodanig te combineren dat de performance optimaal is. Dat EMC daarnaast ook nog eens beschikt over VMWare maakt het alleen maar interessanter. Naast het virtualiseren van hardwareoplossingen kan de managementfunctionaliteit ervan ook worden toegepast op de verkregen MPP-database. Het resultaat van dit alles is dat EMC opeens niet meer (alleen) de storage solution is, maar zich kandidaat stelt als potentiële totaaloplossing voor analyse, Business Intelligence en datawarehousing. De sleutel hiertoe is Greenplum, waarmee het mogelijke effect nog groter zou zijn dan de eveneens recentelijke overname van Sybase door SAP. De gevestigde grootmacht beschikt immers net zomin als Microsoft over eigen hardware, in tegenstelling tot IBM en Oracle dat SUN inlijfde. Ook de cloudoplossingen die EMC door de overname van Greenplum kan aanbieden nemen aanzienlijk toe. Het 'As a Service'-concept kan op de verschillende onderdelen van BI en datawarehousing worden ingezet.

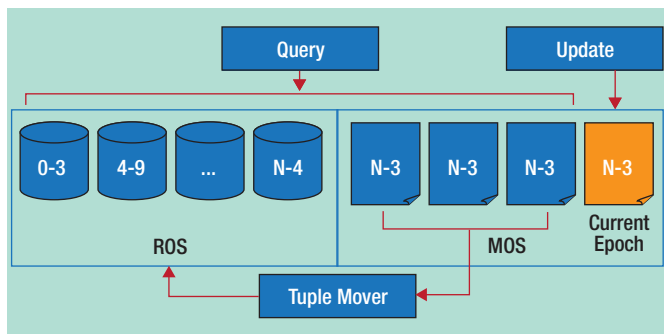
De overname van Greenplum door EMC is nog maar net definitief afgerond. Het beschreven scenario waarbij EMC zich meldt op de markt voor BI en datawarehousing moet nog plaatsvinden, maar lijkt alleszins waarschijnlijk. Of de inkapseling van Greenplum nu ook betekent dat het buiten deze context (als losse database) niet meer wordt ingezet is moeilijk te beantwoorden. Gezien de capaciteiten die Greenplum biedt, heeft het zeker kansen.

Vertica

High performance en hoge beschikbaarheid zijn ook de zaken die Vertica hoog in het vaandel heeft staan. Op basis van research aan het Massachusetts Institute of Technology (MIT) heeft Vertica een column based database ontworpen. Vertica is hiermee momenteel de enige vendor die een MPP-architectuur en een kolomgeoriënteerde database biedt. Het voordeel van een kolomgeoriënteerde database is dat alleen de kolommen die nodig zijn opgehaald hoeven te worden. Daarnaast is in het algemeen de spreiding van waarden in een kolom relatief beperkt, wat voordelen biedt bij datacompressie. Waar Sybase eveneens een kolomgeoriënteerde database heeft (Sybase IQ), maar geen MPP-architectuur en Netezza wel een MPP-architectuur, maar er geen sprake is van een kolomgeoriënteerde database, combineert Vertica de voordelen van beide concepten.

Datacompressie

Behalve MPP en kolomoriëntatie maakt Vertica ook gebruik van datacompressie in de zoektocht naar een goede responsnelheid.



Afbeelding 3: Vertica – Hybride dataopslag (bewerkt naar R.Winter).

Er wordt gebruik gemaakt van verschillende datacompressie-technieken, waarbij gekeken wordt naar datatype, kardinaliteit en sorteervolgorde. De Vertica database kiest automatisch de beste datacompressietechniek voor elke kolom, gebaseerd op een steekproef van de waarden in de kolom. Volgens opgave van Vertica worden compressieratio's van 5 tot 60 keer minder dataopslag bereikt vergeleken met de input. Wat hierbij van groot belang is, is dat bij het gebruik van de compressed data het niet eerst nodig is om de data weer terug te brengen naar de oorspronkelijke staat voordat ze gebruikt kunnen worden. Bij veel databases waarbij ook datacompressie wordt toegepast is dit wel het geval.

De manier waarop data worden opgeslagen is niet in de vorm van tabellen, maar in de vorm van views. Vertica noemt deze views projections. Elke projectie is een combinatie van een aantal kolommen, waarbij de kolommen ook uit verschillende logische tabellen kunnen komen. In dat geval is er dus sprake van een materialized join. Het is uiteraard nodig dat elke kolom tenminste een keer voorkomt in een projectie, zodat alle mogelijke vragen beantwoord kunnen worden. Welke projecties nodig zijn wordt door de Vertica database bepaald die probeert een zo goed mogelijke performance te leveren op basis van geschatte werklast. Al deze automatische beslissingen worden gedaan door de DB Designer. Dat betekent dus ook dat men zich niet hoeft te focussen op hoe de fysieke database eruit moet zien. In plaats daarvan kan alle aandacht uitgaan naar het logische datamodel.

Hybride opslag

Een laatste aspect dat nog genoemd moet worden is het hybride opslagmodel waarvan Vertica gebruik maakt. Alle updates die plaatsvinden worden vastgehouden in de WOS (Write-Optimized Store). Data zijn hier niet-gesorteerd en non-compressed opgeslagen in de volgorde waarin de updates zijn gedaan. De data in de WOS kunnen bevroegd worden. Met een bepaalde regelmaat worden de data vanuit de WOS naar de ROS (Read-Optimized Store) gebracht. Data in de WOS bevinden zich in tijdgebaseerde buckets en worden door de zogenaamd Tuple Mover overgebracht van de WOS naar de ROS. Dit hybride dataopslagmodel zorgt ervoor dat zowel databatches als een trickle-feed ondersteund worden. Volgens Richard Winter, van de Winter Corporation, is het deze combinatie van MPP, kolomoriëntatie en hybride

dataopslag die Vertica onderscheidt van andere databases². De Vertica Analytic Database, nu in zijn vierde versie, is ontworpen door databasegoeroe Michael Stonebraker, bekend van zijn werk aan Ingres en aan de postrelationele database Postgres. Vertica kan worden geïmplementeerd op een cluster van eenvoudige servers. De aanschafprijs is gebaseerd op de 'raw data' die naar het datawarehouse gaan en niet op het aantal processors of gebruikers. Dit alles betekent dat Vertica een relatief goedkope databaseoplossing is. Daarbij geven organisaties die Vertica gebruiken aan dat de implementatie snel is.

Vertica biedt ook een cloudoplossing en was zelfs de eerste database met een cloudoplossing waarbij Amazon Elastic Compute Cloud (EC2) werd gebruikt. Ook Vertica maakt gebruik van MapReduce (van Cloudera) mogelijk, maar heeft ervoor gekozen om dit buiten de database te doen. Vertica draait op Linux, VMWare en in de Amazon Cloud. Het totaal aantal klanten dat momenteel gebruik maakt van Vertica bedraagt meer dan 50. Vertica Analytic Database is onder andere in gebruik bij Sprint, Comcast, Verizon, AT&T Interactive, Blue Cross & Blue Shield.

Conclusie

Aster Data, Greenplum en Vertica bieden allemaal een analytische MPP-database. In de databasemarkt dekt deze categorie databases met name de bovenkant van de markt af, die gekenmerkt wordt door processorintensieve bewerkingen en enorme hoeveelheden data (vanaf 5 Terabyte). Behalve overeenkomsten zijn er echter ook verschillen. Vertica koppelt de MPP-architectuur aan kolomoriëntatie en een hybride dataopslag. Het net door EMC overgenomen Greenplum biedt kolomoriëntatie als optie en biedt daarnaast *self healing fault tolerance*, wat bij inzet van meer dan 100 servers interessant is. Aster Data heeft een native implementatie van MapReduce en een *patent pending* SQL-MapReduce mapping die krachtige analysemogelijkheden binnen handbereik brengt.

Doordat datawarehouses blijven groeien en steeds meer gebruikers bedienen staat de performance steeds vaker onder druk. MPP-databases en de technieken die ze gebruiken worden hierdoor steeds meer mainstream. Adoptie van deze technieken door de mainstream databaseleveranciers of overnames zoals die van Greenplum worden hierdoor waarschijnlijk. Ook voor organisaties in het middensegment is het inmiddels wellicht interessant om de mogelijkheden van analytische MPP-databases verder te verkennen.

Noten

1. 'Magic Quadrant for Data Warehouse Database Management Systems', 28 January 2010, Donald Feinberg & Mark A. Beyer (Gartner).
2. 'The Vertica Analytic Database. Rethinking Data Warehouse Architecture', Richard Winter, 2007-08 (Winter Corporation).

Paul van der Linden (paul.vanderlinden@nl.ibm.com) is senior managing consultant BI/Business Analytics and Optimization en TOGAF8 Certified Architect bij IBM Global Business Services. Dit artikel is op persoonlijke titel geschreven.