

Deeltjesversnellers en databases

Samen op zoek naar subatomaire deeltjes

CERN met zijn deeltjesversneller LHC en zijn wereldwijde Grid is bijna van het eerst uur Oracle-klant. Met behulp van Oracle worden pogingen ondernomen om de fundamentele vragen van het leven te beantwoorden. Omgekeerd heeft CERN Oracle echter ook verder geholpen: CERN is niet alleen een belangrijk alfa- en bètatestter, maar heeft met haar hoge eisen aan performance en schaalbaarheid ook de ontwikkeling van Oracle's techniek versneld.

Het CERN is voor IT-ers veel meer dan een instituut dat onderzoek doet op het gebied van de hoge energiefysica. Het is waarschijnlijk het meest bekend, doordat daar de CERN-wetenschapper Tim Berners-Lee in 1989 een communicatiesysteem ontwikkelde dat wij later internet zijn gaan noemen. Maar ook de 'g' die de Oracle-versies vanaf 10g in hun naam dragen, heeft veel van doen met CERN.

Het CERN is – samen met ondermeer Oracle – in een vroeg stadium gaan experimenteren met gridcomputing. In oktober vorig jaar werd het LHC computing grid officieel geopend. In feite is het Grid een enorme gedistribueerde computer met 100.000 processoren verdeeld over 140 instituten in 33 landen. Meer dan 700 fysici hebben daardoor bijna realtime toegang tot de data van de experimenten en kunnen gebruik maken van de enorme rekenkracht om berekeningen uit te voeren met deze data. De datahoeveelheid is

enorm: per jaar zal de LHC de komende tien tot vijftien jaar ongeveer 15 petabytes (15 miljoen gigabytes) produceren.

Het Grid is maar één onderdeel van het CERN. Voor fysici is de LHC (large hadron collider, oftewel grote hadron botser) veel belangrijker. Het is een deeltjesversneller die onderzoek naar de allerkleinste deeltjes mogelijk maakt. Over veel fundamentele vragen in de fysica bestaat onduidelijkheid en het toetsen van hypothesen daarover is alleen mogelijk door die allerkleinste deeltjes waar te nemen.

Crashdummies

De deeltjesversneller is in feite een soort crash-machine. Twee stralen met deeltjes worden zeventwintig kilometer lang steeds verder versneld door supergeleidende magneten om aan het einde van de rit op elkaar te knallen. Bij de botsingen vallen bekende deeltjes uiteen in kleinere onbekende deeltjes. Er is een theorie die de samenhang tussen de – deels nog niet waargenomen maar wel vermoede – deeltjes verklaart en dat is het zogenaamde standaardmodel (1977). Voor de experimentele bevestiging van die theorie is het nodig het Higgs Bosondeeltje waar te nemen. Uiteindelijk draait het bij CERN om het vinden van antwoorden op vragen als: waaruit is massa ontstaan, wat is de aard van zwarte materie en zijn er verborgen dimensies van de ruimte? CERN is er niet om die vragen te beantwoorden – dat doen



Monica Marinucci, Oracle EMEA R&D Programme Director.

10.000 fysici over de gehele wereld – maar om de experimentele data te leveren op basis waarvan die vragen beantwoord kunnen worden.

Bij de officiële opening van het Grid sprak Optimize met Monica Marinucci (Oracle EMEA R&D Programme Director) en met Dirk Duellmann en Maria Girone, beide fysicus en ondermeer verantwoordelijk voor de Oracle-software en bijbehorende applicaties.

In 1982 was CERN op zoek naar een relationeel DBMS dat de zeer grote hoeveelheid data die het instituut toen al had aan zou kunnen. Na een markton-

derzoek viel de keuze op Oracle 2.3. Oracle breidde zich verder uit naar de algemene IT-infrastructuur en naar de administratieve diensten voor het procurement. Het wereldwijde computergrid van het CERN is van vrij recente datum. Ook daarin speelt Oracle-technologie een belangrijke rol.

Marinucci: "In 2003 begonnen we binnen de samenwerking een zwaarder accent te leggen op grid-technologieën. Dat was de tijd dat Oracle uitkwam met Oracle 10g. We begonnen toen te werken met de Oracle grid-technologie en de distributie van data van tier 0 (CERN) naar tier 1 (aangesloten wetenschappers). CERN heeft meer dan 280 databases geïnstalleerd die samen 600 terabytes aan ruwe data opslaan. Dat zijn dus data gerelateerd aan experimenten maar ook andere zaken, zoals administratieve data."



Maria Girone, één van de fysici die onder meer verantwoordelijk is voor de Oracle-software en bijbehorende applicaties.

CERN gebruikt een mix van 10g en 11g, waarbij de laatste versie steeds verder oprukt. Verder worden producten gebruikt als Oracle RAC, voor de high availability en scalability en Oracle Enterprise Manager, (een tool om gedistribueerde services over het Grid te monitoren en te tunen), Oracle Data Guard voor de back up en disaster recovery en voor het storage management is er Oracle ASM, Oracle automated storage management. Bij al deze software was het CERN betrokken bij alpha- en bètatests en het opstellen van requirements voor nieuwe versies.

Early adopter

Marinucci: "CERN is al die jaren steeds early adopter geweest van onze technologie. Ze werkten samen met onze ontwikkelaars om hun unieke requirements in te brengen, zodat we betere producten konden uitbrengen. We konden onze producten daardoor echter ook valideren en tunen voor deze omgeving die zulke hoge eisen stelt, bijvoorbeeld bij Oracle RAC, Data Guard en Streams. De samenwerking, met name de research binnen Open Lab, geeft ons de mogelijkheid onze technologie naar de bleeding edge (de grens van het technisch mogelijke) te brengen. Tegelijkertijd helpt het CERN aan een gedistribueerde databasestructuur die precies biedt wat ze nodig hebben."

Voor de grote hoeveelheden data die naar de laboratoria buiten CERN moeten worden gerepliceerd, is Oracle Streams uitgekozen. Daarbij ging het niet alleen om replicatie, maar ook en vooral om betrouwbaarheid, availability en schaalbaarheid.

Marinucci: "Ook bij Streams hebben we al in een heel vroeg stadium samengewerkt. Veel van het werk in Open Labs maakt deel uit van het oplossen van zaken met betrekking tot replicatie tussen CERN en tier One sites die soms een slechte netwerkverbinding hebben en soms down gaan gedurende replicatie."

Data type

Dirk Duellmann was direct bij de samenwerking in Open Labs (ook wel make and break program genoemd) betrokken. De feedback van Duellmann voerde zelfs tot de invoering van een nieuw datatype binnen Oracle.

Duellmann: "In het verleden was Oracle een van de weinige databases die geen native ondersteuning had voor de IEEE-getallen in de database. (IEEE staat voor Institute of Electrical and Electronics Engineers, een organisatie, de veel standaarden heeft opgesteld). Door het opslaan en weer lezen van getallen in de database ontstonden kleine afrondingsfouten. Natuurlijk vergt het introduceren van een nieuw basisdatatype in de engine vooral aan de SQL-kant een behoorlijke inspanning, omdat het op het allerlaagste niveau ondersteund moet worden wanneer de data opgeslagen worden in rows. Toen we uitgelegd hadden wat het belang voor ons was, gaf Oracle het een hogere prioriteit en implementeerde het in 10g."

CERN streeft er bij de Oracle-software naar zo gestandaardiseerd mogelijk te zijn. Alleen gedurende de validatiefase zijn er verschillende versies, verder zijn alle database-setups, alle clusters met dezelfde versie van het besturingssysteem van Oracle uitgevoerd. De patches worden op alle clusters tegelijk geïnstalleerd in het ritme van de driemaandelijke patchupdates.

Girone: "Het beheren van een zodanig grote setup is alleen mogelijk doordat we zo focussen op homogeniteit van hardware en van het platform waarop services gedeployed zijn en homogene Oracle-software in diverse versies."

De data die direct verband houden met de experimenten worden niet in Oracle-databases opgeslagen, maar wel alle metadata.

Girone: "De informatie over de status van de detector op het moment dat de botsing plaatsvindt noemen we conditions. Die betreft alle gedetailleerde informatie van de detectors. Het is online



Om de enorme hoeveelheid data die binnenkomt bij het CERN te verwerken is een enorme ruimte vol met servers opgesteld.

processing informatie die nodig is voor de gehele keten van experimenten om de data die uit de botsing ontstaan te begrijpen en te reconstrueren.”

Tot nu toe gebruikte CERN een setup met managed snapshots van een andere database die geüpdatet wordt met een tijdvenster van twee dagen na het origineel en die bovendien gekruist waren. Duermann: “Recent hebben we dit vervangen door Data Guard. We zijn daarmee begonnen bij hardware die op de grens kwam van de garantieperiode. Gezien onze behoefte aan storage zijn we continu bezig nodes, disk en CPU's toe te voegen en aan de andere kant worden dingen eruit gehaald. De invoering van Data Guard betekent in feite dat we alle onderdelen van de high availability architectuur geïnstalleerd hebben en tot dusverre zijn de ervaringen erg positief.” Girone: “De databases hebben nu een fysieke standby in het computercentrum. Daardoor wordt de redundantie van de databases beter, net als de veiligheid. Daarnaast is CERN verbonden met tien laboratoria in de wereld en worden onze data via Oracle Streams dus al gerepliceerd van CERN naar die tien andere sites.”

Door de hoeveelheid metadata betreffende de detectoren en de grote aantal-

len detectoren en gebruikers ontstaat een ingewikkeld geheel. Zo zijn er meer dan duizend gedeployde dataschema's. De applicaties draaien niet op een applicatieserver of databaseserver, maar op de clients.

Duermann: “De totale setup is nogal gecompliceerd, omdat de detectoren van de LHC veel subdetectoren hebben met een zekere granulariteit. Er zijn ook heel veel ontwikkelaars verbonden aan de experimenten. We geven er de voorkeur aan om deze deels niet heel ervaren ontwikkelaars (fysici die het programmeren er als het ware naast doen) toegang te geven tot de data via goed gedefinieerde interfaces. Er zijn wel kleinere projecten die de database direct vanaf SQL-niveau gebruiken. Een ander punt is dat we niet zo heel veel PL/SQL gebruiken, omdat we heel veel gebruikers hebben. Honderden jobs draaien ergens in het Grid, maar voor ons is client side CPU goedkoper dan server side executie. Bovendien zijn er veel analytische functies en floating point verwerking, waarvoor we genoeg client side CPU beschikbaar hebben. De complexiteit van de code is ook erg hoog, dus is het voor ons gemakkelijker om client side applicaties te onderhouden met versioning dan

een hoop PL/SQL versioning binnen de server. Tenslotte is het zo dat de applicaties voortdurend veranderen. De gebruikers leren meer over de detectoren en zullen daardoor ineens waardes gaan gebruiken die ze daarvoor niet gebruikten. Door de applicaties op de clients is het mogelijk de flexibiliteit voor de wetenschappers te behouden zonder een grote belasting voor de dba's.”

Bij alle IT-techniek rond de LHC zou je bijna vergeten dat alles bij het CERN in feite gaat om de zoektocht naar een onvoorstelbaar klein deeltje, waarvan we niet eens weten of het bestaat. Daar komt veel Oracle-techniek aan te pas, maar ook heel erg veel Java. Nu Oracle Sun overneemt wordt CERN nog belangrijker voor Oracle. Voor fysici draait alles om de waarneming van het Higgs Boson. Voor Oracle is het deeltje echter allang zichtbaar. De zichtbaarheid ervan wordt zelfs met de dag groter, zowel door de publiciteitswaarde als door de technische spin-off van de zoektocht.



Dré de Man, tekst en foto's