

Nieuwe versie Caché: afspiegeling van OO, SQL en Java

Datapersistentie is verzekerd

Frans van der Geest

InterSystems Corporation heeft een belangrijke nieuwe versie van Caché uitgebracht. Release 2010 van de object-database moet tegemoet komen aan de vraag naar hoge beschikbaarheid tegen acceptabele kosten. Met deze versie ging tevens de wens in vervulling van Java-ontwikkelaars om een krachtige 'high-volume' dataverwerking te combineren met persistente opslag voor event-verwerkende systemen.

De te verwerken datavolumes nemen toe en daarin krijgt de op Java gebaseerde complex event processing (CEP) met het database-verkeer een groeiend aandeel. Grote database-toepassingen zijn niet meer in staat om aan de vereisten van snelheid en verwerking van een groeiende omvang van data te voldoen. Bijplaatsen van hardware kan enig uitstel betekenen. Maar het eisenpakket blijft groeien en de economische situatie drukt de investeringsbereidheid. Dus na verloop van tijd dienen het prestatieprobleem of een gebrek aan uitbreidingsmogelijkheden (bijvoorbeeld bij nieuwe formaten, JDBC of XML-toevoegingen) zich opnieuw aan.

Tevens gaat de voor het maken van backups benodigde tijd meespelen. De procedures daarvoor worden in een steeds kleiner wordend tijdframe gedrongen. De druk van hoge mate van beschikbaarheid, waarvoor men meestal tot mirroring zijn toevlucht zocht, doet zich bovendien gelden. Dus dat wordt weer schrijven naar schijven. Of, als de opslag met het oog op snelheid in 'solid state' is uitgevoerd, met memory-mapping. Maar dan moeten onvermijdelijk op enig moment ('s nachts bijvoorbeeld) de data uit het RAM naar schijf weggeschreven worden. Ook dat kost tijd en opslag geschiedt veelal in de vorm van 'platte' data, een minder bruikbaar formaat.

Objectoriëntatie vervangt SQL

Het alternatief voor RDBMS'en die de vereiste prestaties of flexibiliteit niet meer halen, zijn de objectgeoriënteerde databases. Een parallelle ontwikkeling naast de opgekomen objectgeoriënteerde programmeertalen, zoals C++, Java, .NET, Perl, Python, Ruby, Smalltalk en Visual Basic.

Inmiddels heeft zich al een flinke gemeenschap gevormd rond de objectgeoriënteerde databases, die een informatieve website (www.odbms.org) onderhoudt. In de beginjaren hadden (na de

eerste hype) objectgeoriënteerde databases een slechte naam en probeerde men die te ontwijken door te spreken over 'postrelationele' databases.

Volgens InterSystems, de ontwikkelaar van database Caché, moeten ook objectdatabases nog tot in lengte van jaren in staat zijn (via SQL) om te gaan met de enorme hoeveelheden data die in meer traditionele vormen worden opgeslagen en bevroegd. Caché is dusdanig ontwikkeld dat het met OO én SQL zeer snel uit de grootste databaseconfiguraties (Terabytes tot Petabytes) informatie kan ophalen en erin wegschrijven. Door de snelheid is het vooral een geschikte databaseoplossing voor applicaties die veel real-time interacties in zich herbergen. We treffen het product in de praktijk dan ook veelvuldig aan bij telecombedrijven, financiële dienstverleners en in de gezondheidszorg. Doch ook in de industrie en bij wetenschappelijke instituten zijn interessante Caché toepassingen gerealiseerd. Een bedrijf als Volvo heeft er zijn data in ondergebracht. De Europese ruimtevaartorganisatie ESA-ESTEC is een grootgebruiker van de database. In Nederland en andere Europese landen zijn door Infor voor bibliotheken omvangrijke projecten gerealiseerd rondom de InterSystems database.

In de nieuwste versie van Caché gaat het niet om 'data mirroring', maar om 'database mirroring'; dit is geen woordenspel, maar een doorwrochte geautomatiseerde oplossing voor geplande en ongeplande downtime. De gespiegelde databases laten zich daarbij ook op twee locaties opslaan en verwerken, zelfs als op beide locaties niet dezelfde configuratie wordt gebruikt. De oplossing is dus veel minder afhankelijk van het (bij updating kwetsbare) besturingssysteem en de clustering van hardware. De kosten van de hoge beschikbaarheidsgraad die InterSystems hiermee bereikt, liggen volgens de ontwikkelaar 30 tot 50 procent lager dan bij concurrerende systemen.

Andere voordelen van de mirroringvoorziening hebben betrekking op configuratiewijzigingen of updates van het besturingssysteem op servers waar één Caché applicatie op draait. Die zijn te activeren zonder merkbare gevolgen voor de beschikbaarheid van de applicatie of de prestaties die in een SLA overeengekomen zijn. Daarnaast is er sprake van een verminderd risico, omdat anders dan bij traditionele replicatieoplossingen met al hun complexe configuraties, Caché voor database mirroring een gemakkelijk te implementeren model gebruikt. En door deze

logische datareplicatie verkleint de mirroring mogelijke risico's, zoals ongelegen updates en verbreiding van fouten die bij fysieke replicatietechnieken in andere systemen kunnen optreden.

Twee toegangen minder beheer

Zoals bekend zijn er veel (database)systemen die soms al vele jaren draaien en waarbij geen behoefte aan vervanging bestaat. Deze databasesystemen en rapportagetools (zoals Crystal Reports) worden vaak wel verbonden met nieuwe functies. De webtoepassingen en webinterfaces zijn daar goede voorbeelden van. Veel applicatiesoftware is al voorbereid op deze vormen van integratie of is er speciaal voor ontwikkeld.

Omdat SQL (in diverse varianten) ouder is dan objectgeoriënteerde databases – de eerste SQL-versie (nog zonder API) werd ontwikkeld in de jaren zeventig – is er een groot aantal applicaties dat van deze opvraagtaal gebruik maakt. InterSystems, van oudsher een karakteristiek technologiebedrijf, heeft zich dan ook veel moeite getroost om in Caché de twee (SQL en OO) samen te voegen in één programma, en er daarbij voor te zorgen dat snelheid en schaalbaarheid gewaarborgd bleven. De beide toegangen zitten diep in het programma verankerd; er is niet gebruik gemaakt van een object-relational mapping tussenlaag als Hibernate. Volgens de makers is de innige samensmelting van SQL met objectoriëntatie uniek. Ook het feit dat Caché beschikbaar is voor Windows, UNIX/Linux/Apple OSX en Open VMS (het past op één CD-ROM en kan op een laptop draaien) toont de weldoordachte opzet aan.

Objectdatabases zijn nog tot in lengte van jaren in staat om te gaan met enorme hoeveelheden data

Bij migratie is in het algemeen weinig tuning noodzakelijk. De producent claimt dat Caché eenvoudiger te beheren is dan andere systemen. Als voorbeeld haalt men aan dat een grote organisatie drie full-time beheerders in dienst had voor de database en dat na de overstap naar Caché er nog slechts één beheerder voor drie dagen per week nodig was.

Virtualisatie met Java-binding

Multiplatform-omgevingen zijn in de meeste organisaties meer regel dan uitzondering. Om in alle omgevingen optimaal functionerende toepassingen te kunnen aanbieden, wordt vaak virtualisatie te hulp geroepen. Ook daar heeft InterSystems naar eigen zeggen zeer doelmatige hulpmiddelen voor. Ontwikkelaars kunnen ervoor kiezen om Java-objecten in de Caché-database op te slaan. Hiermee wordt automatisch een Java-binding gecreëerd en via Java kunnen objecten benaderd worden en ook met SQL is toegang mogelijk. InterSystems heeft hiervoor Caché eXTreme

for Java ontwikkeld. Het betreft een volslagen nieuwe oplossing voor het dilemma waarmee Java-ontwikkelaars geconfronteerd worden bij het bouwen van systemen waar verwerkingssnelheid van essentieel belang is: data veranderen snel en veilige opslag (persistentie) is een absolute eis.

Door directe Java-toegang tot de multidimensionale database engine van Caché mogelijk te maken wordt een spectaculaire prestatiewinst geboekt, die bij praktijktests uitkwam op drie tot zeven maal de snelheid van andere dataverwerkingsoplossingen. De prestaties van Caché eXTreme voor Java zijn vergelijkbaar met die van een in RAM draaiende database, terwijl de persistentie van historische en transactionele data, die vereist is voor CEP en event-gedreven Service Oriented Architecture, verzekerd is. Dit alles zorgt ervoor dat opschaling tot vele duizenden clients probleemloos kan verlopen zonder verlies van prestaties. Een ervaren Java-ontwikkelaar is in staat om binnen een zeer korte leercurve met deze voorziening te werken.

Bij virtualisatie kom je niet om VMware heen. Daar werkt iedere zichzelf respecterende database-ontwikkelaar dan ook mee samen, als het om virtualisatie gaat. De ruime inzetbaarheid van de Caché-oplossingen is volgens InterSystems ook terug te vinden in de vele toepassingen waar de objectdatabase het hart van vormt; naast de meer traditionele toepassingen, zoals gezondheidszorg, bijvoorbeeld ook als serverbasis voor webpagina's, webservices, ODBC en JDBC.

Service, klanten en voorwaarden

Bedrijven en grote organisaties zijn niet op zoek naar de juiste database, maar naar een automatisch systeem dat aan hun wensen voldoet, ook en vooral op langere termijn. In de branche gaat het dus om langere termijn relaties. En belangrijke factoren zijn daarbij de totale exploitatiekosten en de service van de leverancier. InterSystems zorgt er dan ook voor dat elk halfjaar een nieuwe (en tot dusverre steeds snellere) release uitgebracht wordt en dat er verschillende keuzemogelijkheden zijn voor de inzet van de software:

- Huren. Een organisatie huurt bij een serviceverlenend bedrijf een virtuele server op internet; de klant kan Caché huren en daarop installeren. Denk aan grote 'postorderbedrijven'; ook kan Software as a Service een verwante ontwikkeling zijn;
- Kopen. De klant koopt de software en ontwikkelt zelf of met een derde partij de eigen applicatie;
- Revenue share. De klant krijgt Caché in bruikleen, en draagt een deel van de winst af aan InterSystems; is er geen winst dan volgt ook geen winstuitkering.

De objectgeoriënteerde database is nog lang niet op zijn einde. Dagelijks worden meer data vastgelegd en opgevraagd in steeds meer toepassingen. En communicatie wordt daarbij de grote aanjager. Of de technologie het allemaal kan bijhouden is de vraag. Of moeten we het juist andersom zien: of wij de technologie kunnen bijhouden? Het zal altijd wel een wisselwerking blijven.

Frans van der Geest is freelance journalist.