

Klaas Brant over IBM's DB2 Universal Database

Het pannetje van Larry Ellison...

CEO's en topmanagers zijn vaak kleurrijke figuren. Zo ook Janet Perna, general manager Data Management Solutions van IBM. In een van haar toespraken heeft zij Larry Ellison, de CEO van Oracle, wel eens afgeschilderd als iemand die rondloopt met een pannetje. Larry is constant op zoek naar data om die vervolgens stiekem in zijn pannetje op te sluiten.

Het pannetje heet Oracle en wie er eenmaal in zit, komt er nooit meer uit. En wat nu als Larry data tegenkomt die al in een ander pannetje zitten? Dan haalt hij ze eruit en stopt ze alsnog in zijn eigen pannetje...

Een slechte zaak, want zo werkt het natuurlijk vandaag de dag niet meer. Het samengaan van bedrijven is aan de orde van de dag. Iedereen heeft zijn eigen pannetje.

Men heeft geen tijd en geld om data van het ene in het andere pannetje te stoppen. Geen ramp... lekker in het pannetje laten zitten. En dat doet men dan ook heden ten dage. Het converteren van data en applicaties is niet meer aan de orde, zoals dat twintig tot dertig jaar geleden wel het geval was. Het is natuurlijk iets anders als het pannetje gaat lekken en men geen vertrouwen meer heeft in de pannetjesleverancier.

Maar wat nu als we vanuit het ene pannetje data uit het andere pannetje willen hebben? Nou, dan hebben we nog wel wat toeters en bellen om data te kunnen transporteren. Zelfs Larry heeft nog wel het een en ander in zijn keukentje. Mooi toch, of niet? Maar... wat nu als ik data uit het ene pannetje met data uit het andere pannetje wil joinen in één SQL-statement? Ik hoor u al denken... nou, nu ga je het wel heel moeilijk maken. Nee hoor, zeg ik dan, helemaal niet. Deze vraag wordt meer en meer gesteld, en de applicatie-ontwikkelaars krijgen tegenwoordig dezelfde houding als de eindgebruikers van tien, vijftien jaar geleden. "Fix it" in het pannetje, pardon... in de database. Wil ik data joinen uit Oracle en DB2, dan moet dat toch gewoon kunnen zonder replicatie of andere moeilijke zaken? En gelukkig kan dat ook!

IBM heeft een product dat luistert naar de mooie naam Data Joiner. Dit tool, dat is afgeleid van DB2, is in staat zich met allerlei

soorten databases in de wereld te connecten. Het heeft een eigen optimizer, gebaseerd op de starburst optimizer van DB2. In één SQL-statement kan men gegevens uit diverse databases joinen.

Deze optimizer is slim genoeg om zoveel mogelijk werk uit te besteden aan de databases waaraan hij zich connect en de netwerkoverhead van datatransport tot een minimum te beperken. Maar dat is niet het enige. Data Joiner kan meer dan alleen lezen. Ook updates behoren tot de mogelijkheden. Een werkelijk uniek product. Ik ken zelfs mensen die het

gebruiken om complexe SQL te gebruiken op Oracle-data. SQL die Larry's pannetje niet toestaat.

DATA JOINERS TOEKOMST

Nu heeft IBM besloten dat Data Joiner geen apart, stand-alone product hoort te zijn, maar gewoon onderdeel moet zijn van DB2 zelf. Uiteindelijk zal Data Joiner niet meer bestaan. De voordelen zijn duidelijk; op deze manier hoeft men niet twee optimizers te onderhouden en de dba-kennis van SQL-tuning is ook van toepassing op remote (lees niet-DB2-) data. Voor de klanten heeft het als voordeel dat men geen twee producten hoeft te kopen en te installeren. Wat is er nu mogelijk en wat zit eraan te komen?

Op dit moment (DB2 UDB for Unix, Linux en Intel versie 7) heeft men in DB2 de mogelijkheid voor federated database-ondersteuning. Een federated database is een database die zich buiten DB2 bevindt. Dit kan een niet-IBM-database zijn, zoals Oracle of SQL Server - op hetzelfde of op een andere server of ander besturingssysteem. De data zijn volledig transparant beschikbaar voor DB2 in een willekeurig SQL-statement. DB2 zal zoveel mogelijk werk uitbesteden aan de andere server, maar als dat niet mogelijk is de data lokaal halen en dan de SQL lokaal toepassen. Dit houdt in dat ook complexe, alleen door DB2 ondersteunde SQL, zoals recursieve SQL, nu op een willekeurige database toepasbaar is. Maar



ook joins tussen DB2-data en niet-DB2-data zijn mogelijk. *Bye, bye* moeilijke tools en datareplicatie. IBM is hiermee koploper in de markt. Heeft release 7 nog alleen maar read-access op de remote data, op korte termijn zal ook een remote update gedaan kunnen worden vanuit het DB2-systeem. Het systeem dat IBM voor deze federated data gebruikt, is simpel, maar zeer doeltreffend. Door middel van een wrapper wordt de andere server DB2-compatibel gemaakt. Deze wrapper wordt geladen door relational connect, die vereist is om de connectie met de andere database/platform te kunnen maken. De wrapper stuurt de client-software van de andere database aan (bijvoorbeeld Net8 voor Oracle). De client-enabler van de andere database moet dus ook geïnstalleerd worden. Deze wrapper geeft de (on)mogelijkheden van de andere server door aan de optimizer van DB2, zodat deze een goed en snel access path kan maken. Op deze manier is het mogelijk data in het pannetje te laten waar ze zitten. Dat is goed nieuws voor bedrijven die een keuken vol pannetjes hebben. Met DB2 kunnen zij Oracle-data met SQL Server joinen in SQL-constructies die beide databases niet ondersteunen. De performance is natuurlijk afhankelijk van de hoeveelheid data die over het netwerk moeten.¹

ASK DR. SUTOBEUS

Vraag: Het ziet er naartuit dat mijn DB2 geen query's in parallel wil uitvoeren. Waarom?

Antwoord: De configuratieparameter MAX_QUERYDEGREE staat waarschijnlijk nog op de default waarde 0. Door deze parameter aan te passen en de instance te herstarten, is parallel SQL-uit-

voering wel mogelijk. Maar let op, zelfs een monoprocessormachine (normale pc bijvoorbeeld) is in staat SQL in parallel uit te voeren, maar het moet wel mogelijk zijn om dan ook de I/O in parallel te doen. Een standaard IDE-controller kan dit niet, en in dat geval zal het parallelisme niet beter en soms slechter werken. Vandaar de default waarde van 0.

OFFTOPIC

Er gaat een mythe door de wereld dat Java de oplossing is om applicaties portable te maken. Dit is natuurlijk een fabeltje van de eerste orde. Java is een programmeertaal die inderdaad, mits aan bepaalde voorwaarden is voldaan, van het ene platform naar het andere platform overgezet kan worden. De SQL echter in deze Java-programma's zijn geschreven voor een bepaalde databaseserver. Sommige SQL-constructies werken wel op de ene database en niet of zeer slecht op een andere database. Jammer maar helaas, Java's JDBC en SQLJ zijn geen standaards die portabiliteit mogelijk maken. ●

Noot:

1. Wilt u meer weten -en dat wilt u- kijk dan eens rond op www.software.ibm.com/data/db2. In de administration manual van DB2 (zie "library" op de website) vindt u meer informatie. Volgende keer vertel ik meer over de relatie tussen Informix (nu ook van IBM) en DB2.

Klaas Brant (kbrant@kbce.nl) is DB2-specialist en directeur van KBCE. Meer informatie over DB2 is te vinden op www.kbce.nl