

Klaas Brant over IBM's DB2 Universal Database

Meer is beter?

Een trend van de laatste tijd luidt: meer is beter! Waarschijnlijk is dit uit Amerika komen overwaaien, want het is niet iets typisch Europees. Meer winst, meer landen in de EU en meer werk. Meer databases op één afdeling? Probleem of uitdaging, de keuze is aan u. Database-leveranciers spelen hier handig op in met termen als 'zero-maintenance' en 'autonomic computing'. Zijn dit holle kreten of is er meer aan de hand? Bij een recent bezoek aan het IBM lab werd mij duidelijk dat het meer is dan een simpele marketing hype. Neem als voorbeeld Cloudscape, de relationele database die meekwam toen IBM Informix overnam. Janet Perna van IBM gaf eerlijk toe dat ze met het product in haar maag zat. Wat is het en wat moet ik er mee, vroeg zij zich af. Het antwoord op de vraag wat het is, is eenvoudig. Het is een lightweight, zero-maintenance database geschreven in Java; geen file-systeem maar een echte database die 100 procent SQL compatible is met DB2. De database ondersteunt ACID (onder andere commit/rollback). De footprint van 2 MB is zo klein dat hij ook gebruikt kan worden in handheld devices zoals een telefoon. Wat moet IBM met dit juweeltje? Daar heeft Janet zich goed over laten inlichten; weggeven was het advies! Dus Cloudscape is nu open source en heeft onmiddellijk een maatje gevonden in Apache.

Door Apache wordt er een nieuwe naam aan Cloudscape gegeven namelijk Derby. Het wordt druk in open source-land als het gaat om databases (recent doneerde CA ook Ingres). Hoe dan ook, Cloudscape is een goed voorbeeld van een zero-maintenance database.

Maar onze grote databases hoe zit het daarmee? Sturen die ook aan op zero-maintenance? Het antwoord luidt volmondig ja. Nu moet u niet gelijk ontslag nemen als DBA, want voor de grotere installaties zal er op termijn toch nog wel ongeveer 15 tot 25 procent werk overblijven. Wat zijn dan de taken die nu reeds 100 procent automatisch door de database zelf gedaan kunnen worden? Laten we eens inzoomen op DB2 V8.2 for Linux/Unix/Windows (DB2 Stinger). In deze versie zitten diverse componenten die door IBM autonomic Management genoemd worden. Dit is weer heerlijk verwarrend van IBM, want de naam die men eerder gebruikte was SMART (Self-Management and Resource Tuning). Met deze nieuwe features kan de database zelf zijn backup's nemen, zelf de database in optimale conditie houden (auto

online reorg) en zelf besluiten om de statistics bij te werken. Ik durf te wedden dat bij het lezen van dergelijke features diverse mensen beginnen te pruttelen in de trant van "ja maar dat gaat zo maar niet". IBM is zeer serieus en de implementaties in Stinger zijn het gevolg van jarenlang onderzoek. Zo neemt DB2 niet zo maar een backup maar doet dit op een speciale manier. DB2 zal in de gaten houden of de logfiles nog allemaal bestaan en ook de log-files en andere zaken komen op de backup terecht. Ook erg mooi is dat de backup (online natuurlijk) zijn resources optimaal gebruikt. Heeft u de CPU nodig dan doet de backup het zachtjes aan (throttling). Het enige wat u nog moet doen is de backup op een veilige plaats opbergen. En als u remote tape management heeft is dat ook niet eens nodig.



Dat er toekomst zit in al deze technieken blijkt wel uit DB2's nieuwste optimizer LEO (Learning Optimizer). Iedere optimizer gaat de fout in als de gegevens waar de optimizer van uitgaat fout zijn of incompleet zijn. Een voorbeeld: als de verdeling van de data niet homogeen is (bepaalde keys komen extreem veel voor) dan praten we over een *skewed distribution* (schuine verdeling). DB2 kan hiervoor extra gegevens bijhouden die de optimizer helpen met een juiste beslissing te maken. In Stinger is de eerste aanzet gedaan

voor LEO. Indien de database tijdens executie van een SQL statement merkt dat de distributie van de data anders is dan de optimizer inschatte, dan zal er een signaal gegeven worden voor een background runstats om meer en betere runstats te verzamelen. Toekomstige versies van LEO gaan nog een stap verder en gaan zelfs het runstats-proces uitschakelen en leert de database zelf hoe de gegevens er uit zien. Immers, de data zijn er ooit een keer ingekomen en ook updates en deletes zijn bij te houden (real time statistics).

Is meer beter? Als het om autonomic management gaat; wel ja! Een betere database, maar wellicht met minder DBA's. Voor het eerst is niet outsourcing een bedreiging voor de functie van DBA, maar automatisering zelf. Het was te verwachten.

Klaas Brant (kbrant@kbce.nl) is DB2-specialist en directeur van KBCE b.v. Meer informatie over DB2 is te vinden op www.kbce.nl en www.db2-times.com