

VAN DER LANS

Is uw database server wel carefree?



Veel artikelen, lezingen en discussies gaan tegenwoordig over hoe snel bepaalde database servers zijn en hoeveel gegevens ze kunnen beheren. Leveranciers vallen over elkaar om te melden hoe groot de databases zijn die hun producten aankunnen en dat ze immense hoeveelheden query's razendsnel kunnen verwerken. Alles wordt uit de kast gehaald om aan te geven dat ze de grootste en snelste zijn. Maar zijn grootte en snelheid de enige kwalificaties die belangrijk zijn? En trouwens, hoeveel klanten hebben eigenlijk datawarehouses van vijfhonderd Terabytes groot? En wie verwerkt er honderdduizend query's per uur? Uiteraard zullen er organisaties met dit soort eisen zijn, maar de massa heeft dergelijke eisen (gelukkig) niet.

Neem als voorbeeld de TPC-H benchmark resultaten. Leveranciers presenteren altijd trots dat ze weer een record gebroken hebben wat betreft het aantal query's dat ze per uur verwerkt hebben. Maar wat zegt dit nu precies? Wat goed is aan de TPC-H benchmark is dat alleen performancecijfers gepubliceerd mogen worden als daarbij ook aangegeven wordt hoe duur de gehele hardware- en software-opstelling was. De reden is simpel: het is natuurlijk interessant om te weten dat een product honderdduizend query's kan verwerken en dat het concurrerende product in dezelfde tijdspanne slechts de helft kan verwerken, maar als de configuratie van de eerste tien keer zo duur is als de tweede, dan brengt dat de performancecijfers iets meer in perspectief.

Maar worden bij de berekening van die prijs echt alle kosten meegenomen? Hoeveel tijd is er bijvoorbeeld door de specialisten in het laboratorium gestoken in het tunen en optimaliseren van de machines, de disks, en de database server zelf, plus hoeveel tijd heeft het optimaliseren van de opslagstructuur gekost? Hoeveel tijd is gestoken in het nadenken over de ideale indexen en de perfecte wijze van partitionering? En hoe flexibel is het databaseontwerp dat gebruikt is? Ofwel, hoeveel tijd zou er gaan zitten in de aanpassing van dat ontwerp. En dan hebben we het nog niet eens over backups en monitoring. Dit zijn allemaal activiteiten waar in normale projecten veel tijd in gaat zitten. Een databasebeheerder kan er een dagtaak aan hebben en soms zijn complete afdelingen nodig. Tijd is geld en deze kosten zijn helaas niet bij de TPC-H cijfers opgenomen. Dus het is niet helemaal duidelijk hoe duur alles is geweest. Het is zeker interessant om eigenschappen als databasegrootte, query-verwerkingsnelheid en opslagmodel te bestuderen, maar

is het niet net zo belangrijk dat we ook naar eigenschappen als carefreeness en flexibiliteit kijken? Met de mate van carefreeness (zorgeloosheid) geven we aan in hoeverre een database server self-supporting is en met flexibiliteit bedoelen we hoe eenvoudig het is om nieuwe wensen van gebruikers snel te kunnen implementeren.

Als een database server werkelijk honderd procent carefree zou zijn, dan zouden we geen tijd meer hoeven te besteden aan het optimaliseren van de database server, aan het instellen van de bufferparameters, aan het bepalen van welke indexen nodig zijn en hoe de tabellen gepartitioneerd moeten worden. Dit worden lange vakanties voor databasebeheerders.

Het voordeel van een hoge mate van carefreeness is dat veel activiteiten wegvallen, wat een forse hoeveelheid tijd bespaart. En laten we eerlijk zijn, hier besteden we veel meer tijd aan dan we denken. Houd het maar eens bij!

Is totale carefreeness toekomstmuziek? Tegenwoordig bestaan er database servers die het predikaat carefree mogen dragen, zoals illuminate, Netezza en SQLite. SQLite is dan wel niet geschikt om een datawarehouse te ondersteunen, maar de eerste twee wel degelijk. Let wel, ze zijn niet voor honderd procent carefree, maar komen in de buurt. Tevens zal een hoge mate van carefreeness de flexibiliteit verhogen. Voor niet-carefree database servers geldt dat als de behoeftes van gebruikers veranderen en daardoor tabellen en kolommen aangepast moeten worden, de databaseheerder meestal weer opnieuw moet nadenken over de indexen, de parameters en de partities. Als hier niet meer over nagedacht hoeft te worden maar de database server dit zelf automatisch oplost, dan verhoogt dat de snelheid waarmee gebruikersveranderingen geïmplementeerd kunnen worden en dus de flexibiliteit.

Dat database servers grote hoeveelheden gegevens kunnen beheren en dat ze enorme hoeveelheden query's snel kunnen verwerken, is uiteraard belangrijk. Maar laten we ons niet blind staren op deze eigenschappen. De mate van carefreeness heeft een grote invloed op de hoeveelheid kosten die we maken en zou eigenlijk altijd bij aanschaf bestudeerd moeten worden. Het heeft geen zin om een database server aan te schaffen die enorme databases aankan, als we om dat niveau te bereiken een afdeling vol specialisten nodig hebben.

Rick van der Lans is zelfstandig IT-consultant.