

# Automatisch beheer en tuning

**Charles Rozwat over Oracle 10g**

*Op 25 september hield Oracle in Nederland een introductie van de nieuwe versie van het RDBMS. Elders in dit nummer besteed ik daar meer aandacht aan in een aflevering van de reeks 'Event Driven'. Weinigen onder u zal ontgaan zijn dat versie 10g in aantocht is. Voorafgaand aan de presentaties in Amsterdam was ik in de gelegenheid één van de key-note sprekers, Chuck (Charles) Rozwat, te interviewen. Chuck is Executive President Server Technologies bij Oracle. Hij is lid van de Executive Board en rapporteert direct aan Larry Ellison. Chuck is verantwoordelijk voor de ontwikkeling van Oracle 10g.*

Uiteraard staat het gesprek in het teken van deze nieuwe versie van het Oracle RDBMS. De nieuwe versie heeft de g van Grid computing aan zijn naam toegevoegd gekregen. Iets waar Oracle dus veel belang aan hecht.

*Eén van mijn grootste zorgen met betrekking tot Grid computing in database omgevingen is het gegeven dat Grid computing zoals we dat totnogtoe kennen van bijvoorbeeld het SETI-netwerk vóórnamelijk gebaseerd zijn op het delen van rekenkracht. Dat is bij niet-tijdkritische rekenintensieve toepassingen geen probleem. Databases zijn echter vooral I/O gevoelig. Dat laat zich niet zomaar geografisch spreiden. Wat bedoelt Oracle nu met Grid computing in een RDBMS-omgeving?*

Rozwat: 'Ik denk dat je dat juist ziet. Wat wij vooral hebben geprobeerd is het wetenschappelijke grid te scheiden van het commerciële grid. Vergelijk het met de manier waarop het internet is opgezet versus de manier waarop organisaties intranets bouwen. Wij denken dat in een RDBMS-grid wereldwijde spreiding, zoals bij het wetenschappelijke grid, niet belangrijk is. Mensen worden nu in staat gesteld grids te bouwen, en te profiteren van de voordelen, binnen datacenters of zelfs binnen bedrijven. Wat betekent dat nu voor RAC en het Grid? RAC is een belangrijk onderdeel van het Grid. Het maakt clustering binnen de database tier mogelijk. Verder is uiteraard ook clustering binnen de applicatie tier van groot belang. Wij zijn met RAC de enigen die clustering op deze manier aanbieden.

Wij denken dat hetgeen we hebben bereikt met ASM (Automatic Storage Management – red.) ook een belangrijk deel van clustering is. Daarboven staat het beheer van de hele omgeving, en daar speelt Grid-beheer een belangrijke rol, net als het workload management dat we in zowel de application tier als de database tier hebben ingebouwd. Dan zijn er nog de features rondom self-management die de aandacht trekken. Wij denken dat dat van wezenlijk belang is om de schaalbaarheid beheersbaar te houden.'

## Adviezen

*U noemde twee andere punten: ASM en de zelf-beherende en zelf-optimaliserende mogelijkheden van het systeem. Ik heb daar vanuit het verleden wat gemengde gevoelens over. Zo was in mijn optiek één van de grote problemen waar Windows NT mee kampte het feit dat het gebaseerd was op een GUI die iedereen die wel eens Windows for Workgroups 3.11 had geïnstalleerd deed geloven dat installatie van Windows NT een fluitje van een cent moest zijn. Het beheer van een systeem wordt zo eenvoudig dat mensen gaan denken dat er geen kennis nodig is om het systeem te installeren en beheren. De fouten die dan worden gemaakt leiden er vervolgens mede toe dat het systeem als onbetrouwbaar wordt afgeschilderd. Uit het oogpunt van marketing is eenvoudig beheer een pré, maar als beheerder denk ik dat te veel mensen met te weinig kennis problemen zullen veroorzaken.*

Rozwat: 'De manier waarop wij automatisch beheer en automatische tuning benaderen is om het werk van de DBA te vereenvoudigen. Het is niet ingebouwd om DBA's overbodig te maken, of om ongeschoolde medewerkers aan de database te laten werken. We vinden het nog steeds van wezenlijk belang de DBA goed geschoold is. Wat wij proberen te elimineren zijn de basale en foutgevoelige zaken, waar je eigenlijk je tijd niet aan wilt besteden, maar die een Oracle database nu eenmaal van je vraagt. Zo hebben we in de laatste release het hele recovery-proces op de helling gezet. Het was complex, nam veel tijd, maar voegde niet direct wat toe aan het tunen van een systeem. Kijk maar eens goed naar de features voor automatisch beheer. Wij zeggen wel dat het systeem zichzelf auto-

matisch beheert, maar dat doet het niet. Het geeft de DBA adviezen, die de DBA kan raadplegen, waarna hij zelf beslist of hij het opvolgt of niet. Eén van de voorbeelden is bijvoorbeeld ADDM (Automatic Database Diagnostics Monitor, spreek uit als het engels: Adam – red.). Dit rapporteert bijvoorbeeld de slechtste SQL-statements en geeft suggesties om ze te verbeteren. In ons team werkt Graham Wood. Hij is een performance goeroe. Wij hebben deze feature naar hem vernoemd, Graham-Wood-in-a-box.’

## Detailniveau

*U heeft het over ASM, ik weet daar inhoudelijk nog niets vanaf. Ik had ooit een klant die voor opslag had gekozen voor een soort intelligente Raid-oplossing. Het was een combinatie van RAID-10 met RAID-5. Het opslagsysteem optimaliseerde zelf de hot-spots naar het RAID-10 gedeelte. Dit werkte niet: aan het einde van de dagelijkse OLTP-activiteiten stonde precies alle gegevens op RAID-10 die de nachtelijke batches niet nodig hadden. 's Ochtends, na afloop van de batches, was het andersom. Automatisch optimaliseren houdt geen rekening met cyclische verwerkingsslagen, en loopt altijd achter de feiten aan. Heeft ASM daar geen last van?*

Rozwat: 'Ik denk dat het detailniveau waarop de diskvulling wordt beheerd erg klein is. Er wordt gewerkt in eenheden van één MB, waardoor de gegevens voldoende gespreid zullen worden om hotspots te voorkomen. Als we in grotere eenheden zouden werken zouden hotspots kunnen worden bevorderd maar we denken dat we zo hotspots daadwerkelijk helpen voorkomen. Het systeem zal in de loop van de tijd kunnen bepalen wat het

***‘Wij zeggen wel dat het systeem zichzelf automatisch beheert, maar dat doet het niet’***

gebruik van iedere schijf is en op basis daarvan gegevensblokken van de ene naar de andere schijf verplaatsen om zo de belasting gelijkmatig te verdelen. We hebben Graham Wood systemen laten optimaliseren, en daarna hebben we hetzelfde door ASM laten doen. ASM was vrijwel, maar niet helemaal zo goed als Graham.’

*Wat kost ASM, in termen van rekenkracht? Jonathan Lewis zei vorig jaar: Als je een bepaald systeem met Oracle 8i perfect hebt draaien, en je vervangt de database door versie 9i bij gelijkblijvende hardware, dan wordt het systeem 'gratis' 15-20% langzamer. Dit wordt veroorzaakt doordat het dynamische SGA-management extra overhead*

*met zich meebrengt, wat een performanceverlies begrijpelijk maakt. Wat is de invloed van ASM, wat kost het?*

Rozwat: 'We denken dat de extra overhead beperkt blijft tot zo'n twee procent.'

Toen we ons artikel schreven naar aanleiding van de presentatie van Andy Mendelsohn lazen we iets dat een beetje verwarring stichtte. Klopt het dat heterogene clusters kunnen worden ingericht, met bijvoorbeeld een mix van Windows en Linux? Rozwat: 'Nee. Voordat ik naar Oracle kwam werkte ik bij Digital. Daar hadden we al lang DEC clusters toen Unix in opkomst kwam. Eén van de grote vragen toen was of we heterogene clusters mogelijk zouden maken. Onze wedervraag luidde altijd: Waarom zou je dat willen?'

*Je krijgt in ieder geval twee keer zoveel platform-specifieke bugs!*

Rozwat: 'Ja, als mensen heterogene zaken willen opzetten is er altijd het idee van 'zou het niet mooi zijn als we dat ook zouden kunnen', maar er is tot dusver niemand in geslaagd om mij een voorbeeld te geven waaruit het voordeel zou blijken. We ondersteunen het dan ook niet. Binnen een cluster moet tenminste dezelfde architectuur worden gebruikt. Verschillende servers van Dell, HP en IBM binnen één cluster kan, zolang het maar Intel-machines zijn met hetzelfde besturingssysteem, of Windows of Linux. Die twee samen clusteren gaat niet.'

## Replicatie

*Oracle heeft de gegevensopslag met betrekking tot big-endian en little-endian gestandaardiseerd, zodat tablespaces tussen verschillende besturingssystemen en hardware-platforms kunnen worden verplaatst.* Rozwat: 'Als mensen gegevens van het ene naar het andere platform willen verplaatsen moet je dat niet doen door ze samen in een cluster onder te brengen. Transportable tablespaces en de features DataPump en streams zijn bedoeld om gegevens van het ene naar het andere systeem te verplaatsen. Onlangs ben ik nog eens met de hoofdarchitect voor RAC, Angelo Pruscino, om de tafel te gaan zitten om bespreken waarom we eventueel heterogene clusters zouden realiseren.'

*Misschien vanwege de technische uitdaging die ontwikkelaars nu eenmaal graag aangaan?*

Rozwat: 'Ik denk dat je gelijk hebt. Ik ben er niet zeker van dat iemand het daadwerkelijk zou willen gebruiken.'

*Ik zou iedereen adviseren er ver vandaan te blijven.*

Rozwat: 'Ik wil nog even terugkomen op je eerste vraag over geografische spreiding van Grids. Wij hebben een paar RAC clusters draaien waarvan de nodes ongeveer een kilometer uit elkaar staan. Het werkt, en we onderzoeken voortdurend hoe we die afstand kunnen vergroten, maar de lichtsnelheid is daarbij wel een serieus probleem.'

Latency wordt daarbij de bottleneck. Je doorvoersnelheid kan nog zo hoog zijn, bij toenemende afstanden wordt de latency het probleem. Rozwat: 'Wij denken dat mensen datacenters zullen bouwen met bijvoorbeeld Intel/Windows, een ander met Intel/Linux en een derde met Sun/Solaris. Daar kunnen ze clusters en grids bouwen, maar als de verschillende datacenters gekoppeld moeten worden zal dat op basis van replicatie moeten gebeuren, de datacenters zullen niet samen één Grid vormen.'

## Schaalvergroting

*Ik denk dat de licentieprijzen een probleem vormen voor de acceptatie van RAC. Geld besparen op hardware is uiteraard in het belang van Oracle, zo kan meer budget worden vrijgemaakt voor software, bijvoorbeeld van Oracle. RAC-licenties kosten echter zoveel meer dat een twee- of drie-node RAC cluster duurder is dan een SMP systeem met hetzelfde aantal processoren. Gaat Oracle er in de toekomst iets aan doen om ervoor te zorgen dat de instapkosten in RAC-technologie omlaag gaan? (Deze vraag werd 's middags nog onderstreept door een bezoeker, die na de presentatie vroeg of Oracle nog iets gaat doen aan de erg hoge kosten die zijn gemoeid met de high speed internode connect).*

Rozwat: 'Als je kijkt naar wat we in de loop van de jaren hebben gedaan, dan hebben we meer en meer in het basisproduct gestopt in plaats van het hanteren van talloze opties, die voor veel geld moeten worden aangeschaft. RAC is twee jaar geleden uitgebracht. Wij zien het als een zodanig uniek product dat we verwachten dat men daar geld voor over heeft. Uiteindelijk heeft Oracle ook aandeelhouders, en willen we als bedrijf winst maken. We denken dat mensen de waarde ervan inzien, er zijn enorme besparingen op hardware mogelijk, en ook de verhoogde beschikbaarheid heeft z'n waarde voor de gebruikers. Amazon bijvoorbeeld heeft berekend dat iedere minuut uitval rond de drie miljoen dollar kost. We denken dat RAC extra waarde heeft voor de klant. Op de lange termijn willen we dat meer mensen deze technologie gaan gebruiken. We hebben voor deze tournee presentaties over functionaliteit van 10g bewust van het licentiebeleid gescheiden. We zullen het licentiemodel bekend maken voordat de producten eind dit jaar uitgebracht worden, maar we willen die discussie scheiden. Ik kan je in algemene zin vertellen dat we willen dat RAC door veel mensen gebruikt zal gaan worden en dat ons doel op lange termijn is dat meer features beschikbaar komen voor meer mensen. Dit is een algemene richting, het is geen uitspraak over wat er al dan niet in een bepaalde licentie wordt opgenomen.'

*Het grid is leuk voor grote datacenters, en dat geldt voor meer nieuwe features. Neem nu bijvoorbeeld een gemeente, of een middelgroot bedrijf, met drie servers en vijftien databases. Welke features bevat 10g die hen moet laten besluiten tot een upgrade? Rozwat: 'Het belangrijkste is dat we ervoor hebben gezorgd dat de migratie goed werkt. Als je nog steeds op 8i werkt kan*

je direct upgraden naar 10g, dat hoeft niet via 9i. Verder zit er veel interessants in, zoals bijvoorbeeld performance. PL/SQL is dramatisch versneld. Systemen die veelvuldig PL/SQL gebruiken zullen hiervan veel voordeel ondervinden. Verder is de performance over de hele linie verbeterd. Eén van de doelen die we ons stelden bij de bouw van deze versie was om de performance letterlijk te verdubbelen. Dat betekent dat een klant die nu met zijn hardware aan de grenzen van zijn systeem zit, door een upgrade nog een tijd met de bestaande hardware kan door-

***'We onderzoeken voortdurend hoe we de afstand tussen nodes kunnen vergroten, maar de lichtsnelheid blijft wel een serieus probleem'***

werken. Andere klanten kunnen misschien zelfs hardware buiten gebruik stellen en daarmee op kosten van onderhoud en beheer besparen. Verder is ook de automatisering van opslag, tuning en beheer voor de kleinere klant interessant.'

## Previews

Helaas liet de tijd het mij niet toe op dit laatste wat dieper in te gaan. De mogelijkheden en voordelen die Oracle claimt zijn enorm, en ik ben erg nieuwsgierig naar de eerste previews van deze nieuwe release. Er zal opnieuw flink moeten worden bijgeschoold om al deze kennis te beheersen. Als het waar is dat er steeds meer vervelend werk vervalt, kan het niet anders dan dat er ook DBA-banen komen te vervallen. De besparingen als gevolg van de schaalvergroting zullen ergens vandaan moeten komen. Als dat allemaal doorgang vindt zal ook Oracle University dit gaan merken, doordat er uiteindelijk minder DBA's hun cursussen komen volgen. Oracle belooft een interessant begin van het nieuwe jaar!

## Carel-Jan Engel

werkt als onafhankelijk Oracle Consultant. Hij is sinds 1985 in verschillende rollen bezig met ontwikkeling en beheer van systemen op basis van Oracle. Hij ondersteunt organisaties van uiteenlopende omvang in veelal kortlopende trajecten rondom inrichting, tuning en troubleshooting van Oracle databases (e-mail: cjpengel.dbalert@xs4all.nl).