

Goed gereedschap is het halve werk

Gebruiksimpresie van Oracle DBA-tool (8)

Rondom het Oracle RDBMS leveren diverse leveranciers hulpmiddelen die het leven, maar vooral het werk van een DBA helpen te veraangenamen. Een eerste blik in de Oracle Toolgids in het decembernummer van Optimize leert dat er tientallen tools worden vermeld die op één of andere manier onder deze definitie vallen. Daarnaast zijn er ongetwijfeld nog tools die een bespreking waard zijn, maar die niet in Rubriek 5 (databasebeheer) van de Oracle Toolgids staan, al is het alleen maar omdat blijkt dat DBA's soms voor ontwikkelaars bedoelde producten graag 'misbruiken' voor het beheer.

Performance-tuning rondom de Oracle database is, ik heb het al vaker geschreven, vaak omgeven met veel magie. Probeer dit eens, wijzig die parameter, database herstarten: hoe gaat het nu? Al met al een erg onbevredigende manier van werken, niet in het minst voor de gebruiker die na de herstart maar al te vaak met een langzamere in plaats van snellere database wordt geconfronteerd. Eigenlijk wil je dat performance-tuning een voorspelbaar karakter krijgt. Zeker in een 24x7 omgeving, waar een herstart van de database echt niet zomaar even tussendoor kan. Het allermooist is het als de ontwikkelaars tijdens de bouw van de applicatie een optimale performance meteen kunnen meenemen.

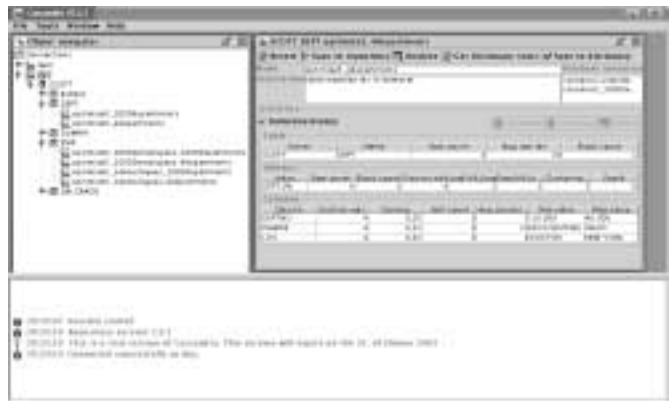
Verantwoording

Deze serie tool-besprekingen is gebaseerd op mijn persoonlijke ervaringen met de betreffende tools. Ieder besproken tool wordt geïnstalleerd in de omgeving waarvoor het is bedoeld en de software wordt met de nodige testcases uitgeprobeerd. Het artikel kan worden gezien als een 'gebruiks-impresie'. De weergave van de resultaten is daarom per definitie subjectief en voor discussie vatbaar. Ik stel me graag open voor discussie. Opmerkingen, vragen en suggesties (ook voor eventueel in volgende afleveringen te bespreken tools) zijn welkom op mijn email-adres. Op uw reacties kom ik graag terug.

Statistieken

De cost-based optimizer binnen het Oracle RDBMS brengt de vervulling van een aantal van deze wensen binnen handbereik. Een belangrijk kenmerk van de cost-based optimizer is dat deze zijn aanvalsplan voor een bepaald SQL-statement opstelt op basis van eerder verzamelde statistieken. Het resultaat van vers ontwikkeld SQL wijkt dan ook in een productieomgeving vaak dramatisch af van de door de programmeur verwachte resultaten. Immers, de hoeveelheden data in de testomgeving zijn aanzienlijk kleiner dan in de productieomgeving, en daarmee kloppen ook de statistieken niet meer.

Dit lijkt een nadeel, maar laat zich met wat werk in een voordeel omzetten. Het is mogelijk om de statistieken van de productieomgeving te transporteren naar de ontwikkel- of testomgeving. Het gevolg van het overzetten van de statistieken is dat ook in de ontwikkel- of testomgeving de cost-based



Afbeelding 1. Het openingsscherm van Cassandra.

optimizer uitgaat van de in de productieomgeving geldende aantallen, en dientengevolge met een identiek executie-plan komt. Kortom, op deze manier kan het tunen van SQL op een voorspelbare manier worden gedaan, terwijl de productieomgeving met rust wordt gelaten. Uiteraard werkt de rule-based optimizer niet op deze manier, die was naar zijn aard al voor-

spelbaar. Net zo voorspelbaar is zijn einde: in Oracle 10g is er geen rule-based optimizer meer. Als voor een SQL-statement benodigde statistieken nog niet bestaan, worden ze tijdens het parsen van het SQL-statement alsnog verzameld. Of je daarmee blij moet zijn kan ik nog niet beoordelen, maar vaststaat dat statistieken omtrent tabel- en indexvulling steeds belangrijker worden in het tuning-proces.

Cassandra

Nu heeft het beheren van statistieken nogal wat voeten in de aarde, maar het is te doen. Dat moet beter kunnen, dachten ze bij ABASE Consulting in Helsinki, Finland, en zij bouwden er een tool voor: Cassandra. In deze aflevering bespreek ik dit tool.

Historie

Cassandra is nog niet zo oud, met de ontwikkeling werd in 2001 begonnen. De eerste versie werd tijdens het jaarlijkse evenement van de Finse Oracle User Group aangekondigd. Aanleiding voor de ontwikkeling was de groeiende frustratie over steeds meer klachten in de trant van: 'Ik heb mijn systeem in productie gebracht, en nu performt het niet meer'. Was het oorspronkelijk een idee van Thomas Atila, inmiddels werkt er een team van vier mensen regelmatig aan de verbeteringen van het tool.

Productopbouw

Cassandra bestaat uit één programma, geleverd in een .jar file. Met de lokaal beschikbare java runtime kan het programma worden uitgevoerd. Cassandra maakt gebruik van de JDBC thin client of van de ODBC driver. Daarmee kan Cassandra worden



Afbeelding 2. Het venster met het tabel scenario.

gebruikt op de diverse Windows smaken en op alle Linux/Unix varianten. Voorwaarde is dat er een Java Runtime Environment versie 1.3.1 of hoger aanwezig is. Het programma bestaat uit een client- en een servergedeelte. Het servergedeelte is als (wrapped) packaged procedures uitgevoerd. Bij het programma zit een redelijk uitgebreide handleiding van 68 pagina's. Deze handleiding wordt geleverd in PDF formaat.

Cassandra maakt gebruik van de JDBC thin client of de ODBC driver, en kan worden gebruikt op de diverse Windows- en Linux/Unix varianten

Het verplaatsen van statistiekgegevens van de ene naar de andere database is helaas (nog?) niet mogelijk vanuit de GUI. Wel worden daarvoor scripts meegeleverd, waarmee dit (met behulp van de Oracle tools SQL*Loader en SQL*Plus) mogelijk wordt gemaakt. Geëxporteerd worden daarbij de actieve statistieken in een database, geïmporteerd wordt er in de Cassandra repository.

Installatie

Geen wizard of installatieprogramma, maar eenvoudigweg de programma-directory kopiëren van de CD naar de gewenste programmadirectory. Ook een java runtime voor Windows wordt op de CD meegeleverd, evenals de JDBC drivers voor Oracle. Na installatie van deze componenten moet een hyperlink worden aangepast aan de directory-namen die bij de installatie zijn gekozen. Het is allemaal net even lastiger dan een simpele druk op de knop, maar mag geen problemen opleveren.

Nadat de client-software is geïnstalleerd moet in een configuratiebestand worden vastgelegd welke databases kunnen worden benaderd, welke schema's daarvan moeten worden onderdrukt (SYS, SYSTEM etc.) en welke schema's mogen worden getoond. De configuratie hiervan gaat prima met willekeurig welke tekst-editor, maar laat de gebruiker wel veel mogelijkheden invoerfouten te maken. Ik had hier liever gezien dat de configuratie via het menu vanuit het tool was gerealiseerd.

Tenslotte moet er aan de database kant ook nog het een en ander worden gedaan: er moet een Cassandra repository worden aangemaakt (standaard onder de user ABASE) en binnen deze repository moeten achtereenvolgens een aantal packages en een plan_table worden aangemaakt. Deze laatste wordt gebruikt om resultaten van een explain plan ('toegangspadanalyse') te bewaren.

Vullingsgraad

Het programma opent met een browser, waarin de bij installatie geconfigureerde databases worden getoond. Zodra op een database wordt geklikt verschijnt een login-venster om in te loggen bij de Oracle database. Na het inloggen worden de eerder geconfigureerde schema's getoond.

Van het schema worden de daarbinnen gedefinieerde tabellen getoond. Het beheer, en het 'spelen met' de verschillende statistieken wordt vervolgens uitgevoerd met behulp van zgn. scenario's. Dat meteen de belangrijkste eigenschap van Cassandra: van iedere tabel kunnen scenario's worden gemaakt. Het scenario van een tabel dat is geactiveerd wordt in de object browser gemarkeerd met een rood driehoekje.

Een scenario omvat de statistische eigenschappen van een tabel en zijn indexen. Scenario's kunnen een naam krijgen, en ieder scenario kan afzonderlijk worden geactiveerd door het in de Oracle data dictionary te laden. Op deze manier kan de cost based optimizer worden gevoed met statistieken over tabelgroottes die in werkelijkheid helemaal niet bestaan. Hiermee kan nauwkeurig worden voorspeld wat de kosten van query's in de toekomst zullen zijn, als de database zijn uiteindelijke vullingsgraad bereikt. Het programma heeft een help-knop, maar daaronder wordt allen programma-informatie verstrekt.

Tabel scenario's

In eerste aanleg is er uiteraard nog geen informatie beschikbaar. Met een rechter muisklik wordt een nieuw scenario aangemaakt, dat vervolgens naar wens kan worden gevuld: met zelfgekozen waarden, vanuit de Oracle data dictionary of door de tabel opnieuw te analyseren.

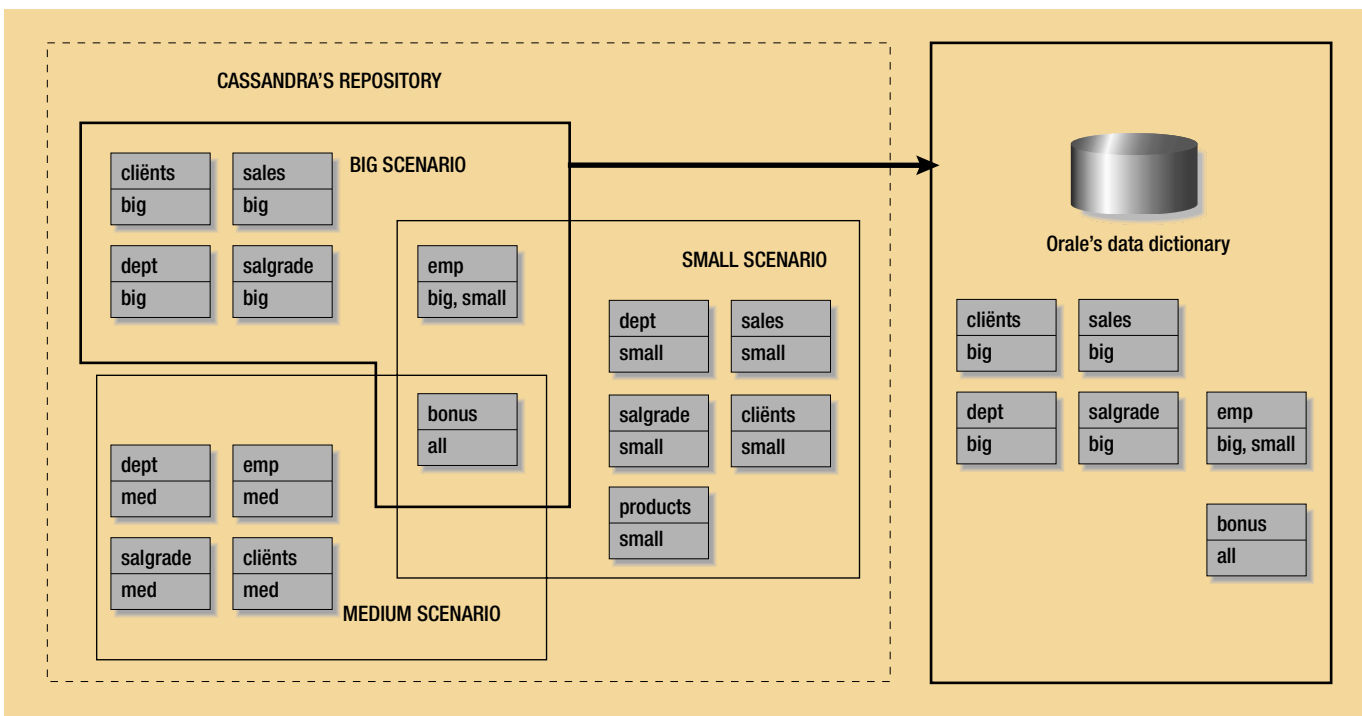
De documentatie verklaart de verschillende velden binnen de statistieken. Achtergrondkennis van de gegevensopslag binnen

Oracle is wel vereist: het is zonder meer mogelijk de cost based optimizer te laten geloven dat er in één datablock 1,5 miljard rows van een tabel zijn opgeslagen. Dan geldt de uitspraak van één van mijn AMBI-docenten vroeger: 'Results are unpredictable.....'

Terecht waarschuwt de documentatie dan ook uitgebreid voor de risico's van het gebruik in een productie-omgeving. Als de statistieken worden geladen, heeft dit invloed op de performance van alle SQL dat de betreffende tabellen benadert. Ongelukkig gekozen statistieken kunnen de performance dan ook dramatisch verslechteren. Uiteraard is er niets op tegen om de statistieken in een productie-omgeving te verzamelen. Dat kan door tabellen te analyseren, maar Cassandra kan ook de reeds aanwezige statistieken in tabelscenario's opslaan.

Database scenario's

Een krachtig hulpmiddel is de mogelijkheid verschillende tabelscenario's te combineren in database scenario's. In feite is een database scenario niet meer dan een verzameling individuele tabelscenario's. Eén tabelscenario kan deel uitmaken van meerdere database scenario's. Verschillende domeintabellen hebben immers een vaste omvang, ongeacht de grootte van de andere tabellen. Waar de andere tabellen een scenario per geschatte databaseomvang hebben, is voor een domeintabel één scenario voldoende. De hele verzameling tabelscenario's die is gegroepeerd in één database scenario kan met één druk op de knop in de Oracle data dictionary worden geladen.



Afbeelding 3: Het verband tussen Database scenario's, Tabel scenario's en de Oracle datadictionary.



Afbeelding 4: Het Explain Plan tool

De database scenario's worden benaderd met een rechter muis-klik op de database in de objectbrowser. Dat vind ik onlogisch: van de tabellen worden ze netjes in de browser getoond, en van de database moeten ze via een menuutje worden benaderd.

Explain Plan

Met het 'Explain Plan' tool kan uiteindelijk snel worden beoordeeld wat de effecten van een bepaalde statistiek zijn op de SQL-statements. De verschillende uitkomsten kunnen (met omschrijving) worden opgeslagen, zodat bij gestructureerd gebruik van het geheel precies kan worden vastgelegd welke effecten de verschillende statistieken op de diverse SQL-statements hebben.

De SQL-statements moeten worden ingetypt of met knippen-en-plakken vanuit een ander programma worden overgesleept.

Productgegevens

Producten	Cassandra 2.5.1 (evaluatieversie).
Leverancier	Abase Consulting AB Helsinki, Finland (www.abase.fi)
Prijs	£ 5.950,00 voor één named user. Overige prijzen op aanvraag.
Onderhoud	(incl. helpdesk en nieuwe versies): 15% per jaar.
Positief	<ul style="list-style-type: none"> • Simpel, rechttoe-rechtaan statistiekbeheer. • Speelmogelijkheden met verschillende scenario's geven snel inzicht in effect van verschillende statistische waarden • Eenvoudige intuïtieve gebruikers - interface
Negatief	<ul style="list-style-type: none"> • Nog ietwat hoekige installatieprocedure • Onlogische plaats van de database scenario's in de object browser • Geen online-help beschikbaar

Het grootste probleem vind ik dat alle tabelnamen moeten worden gespecificeerd inclusief de betreffende schema-naam: Cassandra draait binnen het schema van de Cassandra-repository, en dat is meestal niet het schema van de applicatie. Het zou praktischer zijn geweest als het 'Explain Plan' tool apart zou kunnen inloggen in het schema van de applicatie-eigenaar of -gebruiker. Dat zou wel weer betekenen dat deze gebruiker toegang moet krijgen tot (een deel van) de Cassandra-repository, maar dat vind ik een minder groot probleem.

Samenvatting

Mooi door z'n eenvoud, is mijn eerste associatie als ik mijn mening over Cassandra moet samenvatten. Het is duidelijk een product dat nog niet al te gek lang op de markt is, er is nog een aantal wensen over. Met name de setup kan beter, al mag dat voor degene die op dit niveau met tuning bezig is geen problemen opleveren. Toch zie ik nu al veel gebruiksmogelijk-

Als de statistieken worden geladen, heeft dit invloed op de performance van alle SQL dat de betreffende tabellen benadert

heden, die helpen gecontroleerd en voorspelbaar te tunen. Met name in omgevingen waarin op grote productiedatabases regelmatig nieuwe software-releases worden geïnstalleerd kan goed Statisticsmanagement de ontwikkelaar helpen SQL te produceren dat van meet af aan naar verwachting presteert. De prijs is stevig, maar voor de (ontwikkeld-)DBA die dit op zijn waarde weet te schatten is het niet te duur.

Carel-Jan Engel

werkt als onafhankelijk Oracle Consultant en ondersteunt organisaties van uiteenlopende omvang in veelal kortlopende trajecten rondom inrichting, tuning en troubleshooting van Oracle databases.

E-mail: cjengel.dbalert@xs4all.nl