

Benchmark in SQL Server-omgeving

# Test met Quest bevestigt gunstig beeld prijs/prestatie

Bram Dons

**B**enchmark Factory van Quest Software staat in dit artikel centraal. Een benchmark is een softwaretool dat een consistente methode biedt voor het doorlichten van een systeem op eigenschappen als stabiliteit, snelheid en betrouwbaarheid. Bram Dons bespreekt Benchmark Factory, en voert hiermee een TPC-C belastbaarheidstest uit op een SQL Server 2000 Enterprise database-omgeving.

Benchmark Factory biedt een accurate methode voor het meten van het aantal transactions per second (tps) en andere

kenmerken van 'systeem-onder-test'. Naast een aantal standaardtests biedt Benchmark Factory de ontwikkelaar van client/server-systemen de mogelijkheid zelf benchmarks te definiëren. Het tool ondersteunt verschillende serversystemen, waaronder database-, file-, transaction-, cache- en Internet-servers.

De software bestaat uit drie basiselementen: Visual Control Center, Agent en Repository. Deze componenten draaien als multi-threaded Windows-applicaties op een Windows NT/2000-platform (zie figuur 1). Visual Control Center is het 'dashboard' van de Benchmark Factory-test-omgeving. De agents draaien op de werk-

stations, die requests naar het te testen systeem verstuurt. Agents zijn ook multi-threaded applicaties, die op een enkel werkstation meer dan duizend gebruikers kunnen simuleren. Benchmark Factory is in theorie oneindig schaalbaar: horizontaal door de ondersteuning van honderden werkstations (waarop Agent-software draait), verticaal door op iedere agent duizenden threads te draaien.

De Repository is de opbergplaats voor de gemeten resultaten. Default maakt hij gebruik van de meegeïnstalleerde SQL Anywhere-database van Sybase. De repository wordt op hetzelfde systeem geïnstalleerd als waarop Visual Control Center draait. Naar wens kan ook een SQL Server-database worden gebruikt.

Benchmark Factory verzamelt een vaste hoeveelheid statistische gegevens, zoals de algemene doorvoer (gemeten in transacties per seconde/minuut of het aantal verzonden bytes) en gedetailleerde informatie over een individueel werkstation uitgevoerde transacties. De statistische gegevens dienen naderhand voor analyse, om de capaciteit van een systeem te kunnen voorspellen.

## SCHAALBAARHEID

Benchmark Factory is geschikt voor het testen van de belasting en uitspraken te kunnen doen over de mogelijkheid om het systeem te kunnen uitbreiden. Doordat het tool in hoge mate schaalbaar is, is het uitermate geschikt voor capaciteitsplanning en het optimaliseren van de systeempres-

## Wat is een benchmark?

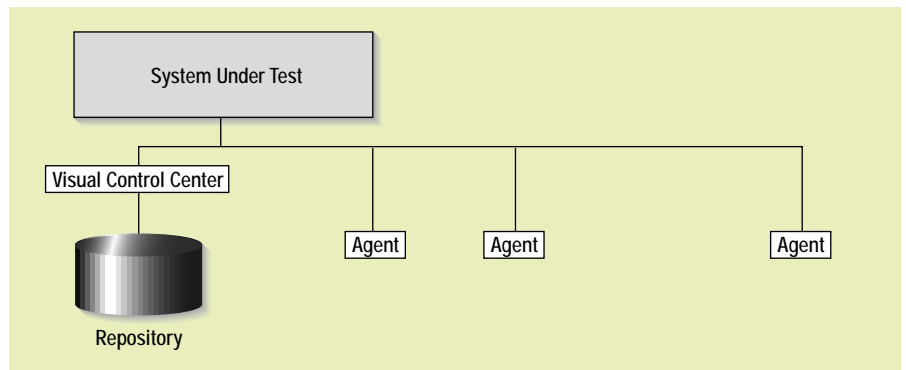
Een benchmark is een methode voor het meten en vergelijken van twee systemen. Het is een softwaretool dat een bepaalde applicatiebelasting op consistente, herhaalbare wijze simuleert, zoals in een praktijksituatie. De software licht de systeemprestaties door en vergelijkt deze naderhand met de meetresultaten van andere systemen. Het testmodel van de benchmark kan uit een aantal applicaties zowel software als hardware evalueren. Zo'n model werkt beter dan de eigenlijke applicatie, omdat het statisch van aard is en reproduceerbare resultaten oplevert. De applicatie bevat te veel variabelen die de resultaten bij het draaien van meerdere sessies kunnen beïnvloeden. Een benchmarktesttool waarborgt consistente resultaten doordat het aantal variabelen in de testomgeving zoveel mogelijk wordt beperkt. Door bovendien de testomgeving af te schermen van invloeden van buitenaf, is de kwaliteit van de test verder te verbeteren.

Benchmarks worden voor veel doeleinden gebruikt: productvergelijkingen, optimalisatie van systeemprestaties, capaciteitsplanning, productontwikkeling en meer. De meest bekende in de IT zijn de standaard benchmarks AS3AP, die van de Transaction Processing Council (TPC-C en -D), NetBench, ServerBench, WinBench van Ziff-Davis Publishing Suite en SGI's WebStone-benchmark. De genoemde benchmarks zijn niet ontwikkeld om het laatste antwoord te geven op de vraag in hoeverre producten de door leveranciers geclaimde prestaties in de praktijk zullen waarmaken. Veeleer bieden zij gebruikers een handvat om producten op een realistische manier met elkaar te vergelijken.

taties. Door de belasting van duizenden gebruikers op een database-, file-, internet- en messagingserver na te bootsen, kan de werking van een systeem onder stress in praktijk nauwkeurig worden nagebootst. Tests kunnen draaien op zowel stand-alone systemen als complexe clusterservers. De 'horizontale' schaalbaarheid wordt bereikt door de ondersteuning van honderden werkstations; 'verticaal' kan de software op iedere agent honderd tot duizend gebruikers simuleren. Vanuit één enkel workstation kunnen duizenden gebruikers worden gesimuleerd, gekoppeld aan de mogelijkheid complexe *sequences of events* door de gebruikers zelf te definiëren en laten uitvoeren.

**VISUAL CONTROL CENTER EN AGENT SCREENS**

Benchmark Factory wordt vanuit het Visual Control Center 'home screen' bestuurd, van waaruit bijna alle benchmarktaken zijn uit te voeren. Via het Visual Control Center kan een benchmark gecreëerd, gemodificeerd en uitgevoerd worden en zijn de gemeten testresultaten in verschillende contexten en vormen zichtbaar te maken. Het linkervenster laat de boomstructuur van een bepaald 'project' zien. Een project kan verschillende testonderdelen bevatten, in combinatie met een of meer benchmarks. Op basis van Available Parts wordt een Assembled Benchmark samengesteld. Available Parts kunnen bestaan uit enkel- of meervoudige



**FIGUUR 1: COMPONENTEN BENCHMARK FACTORY.**

gebruiker-Scripts of externe functies. In het rechternvenster kan men kiezen een viertal tabbladen: Run Queue, Agent Status, Summary en Real-Time Summary. Het Run Queue-tabblad is de plaats van waaruit de benchmarktaken worden voorbereid en uitgevoerd. Agent Status geeft de status van een bepaalde agent weer. Een agent wordt op een fysiek clientsys-

*Benchmark Factory biedt een accurate methode voor tps-meting en andere karakteristieken van een 'systeem-onder-test'*

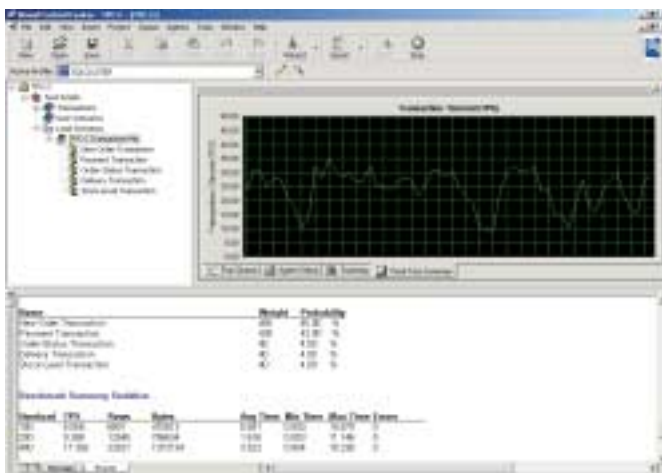
teem geïnstalleerd, waarop meerdere virtuele gebruikerssessies kunnen worden uitgevoerd.

Na het draaien van de benchmark staan de gebruiker talrijke grafieken en gegevens ter beschikking. In de Summary Graphs-sectie is een aantal grafieken te

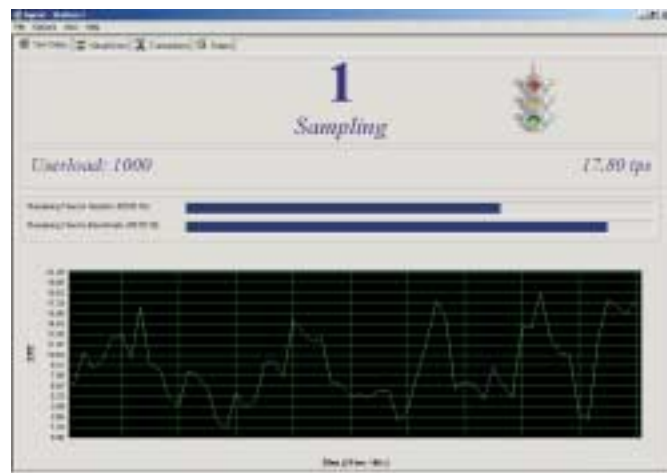
zien die onder meer gebruikersbelasting, TPC en responstijden laten zien. De Results-map toont de statistieken voor iedere gebruikersbelasting. Deze informatie is via het clipboard naar bijvoorbeeld Excel te kopiëren. De General Statistics-map toont algemene informatie, waaronder tps en BPS. Bovendien toont het minimum, maximum en gemiddelde transactietijden. De Users-map toont onder meer gedetailleerde informatie over elk van de gesimuleerde gebruikers, waaronder User ID, tps, aantal uitvoeringen, aantal geretourneerde databaserijen, aantal bytes en statistieken over transactietijden. Distribution en Summary Graph tonen gegevens per virtuele gebruiker, zoals Transaction Time Distribution, Response Time en Latency Distributions.

**INSTALLATIE**

Installatie kan als Standard- of Lab-configuratie. De eerste is voornamelijk bedoeld



**FIGUUR 2: VISUAL CONTROL CENTER MET REAL-TIME SUMMARY.**



**FIGUUR 3: AGENT SCREEN.**

voor database-evaluatie en het testen van systemen zonder netwerk. De vanuit het setup-menu te kiezen Standard Installation installeert Agent en Visual Control Center op hetzelfde systeem. Lab-installatie is bedoeld voor testomgevingen waarin meerdere werkstations voor de belasting moeten zorgen. Via deze optie wordt allereerst Visual Control Center op één workstation geïnstalleerd en vervolgens via het netwerk -via de gedeelde directory NetInstall- op alle agent-systemen de agent-software.

Voor Visual Control Center is minimaal een Pentium II 160 MHz-systeem met 32 MB RAM en 100 MB vrije diskruimte nodig, waarop Windows NT/2000 draait; Windows 2000 wordt aanbevolen. Voor een Agent-systeem volstaat 100 MHz met 16 MB.

Benchmark Factory maakt als 'repository' van Sybase's SQL Anywhere v6.0-database gebruik om testscripts en -resultaten op te slaan. Standaard levert Benchmark Factory een aantal testomgevingen mee, voor een database-, web-, file- en messa-

ging-omgeving. Hierna bespreken we een TPC-C database-test met de SQL Server 2000 Enterprise Edition.

**'REAL WORLD LATENCIES'**

Benchmark Factory kent talrijke benchmarkopties. Voor de database-omgeving is een aantal opties van speciaal belang. Om zoveel mogelijk het werkelijke gedrag van databasegebruikers te kunnen nabootsen, hanteert Benchmark Factory het concept *real world latencies*. Hiermee wordt de tijdsduur gedefinieerd die de benchmark wacht bij het doen uitgaan van een request naar de databaseserver. De praktijk: een doorsnee databasegebruiker voert een SQL-query uit, ontvangt de resultaten daarvan en denkt een moment na voordat hij de volgende query uitvoert; de benodigde denktijd kan ook worden besteed aan het invoeren van een nieuwe opdracht, analyse van de getoonde resultaten of koffie drinken. Een benchmark uitgevoerd zonder latency's, waarbij transacties zo snel mogelijk worden uitgevoerd,

belasten de databaseserver namelijk aanmerkelijk minder dan in de praktijk het geval zou zijn.

Benchmark Factory biedt diverse opties voor de instelling van latency's in een multi-user testomgeving: no delay, Inter-Arrival Time, Keying Time en Think Time. Afgezien van no delay, biedt iedere andere latency-instelling opties als: absolute, uniform (binnen een bepaald bereik), normal (op basis van een gemiddelde waarde) of negative exponential of Poisson. Alhoewel

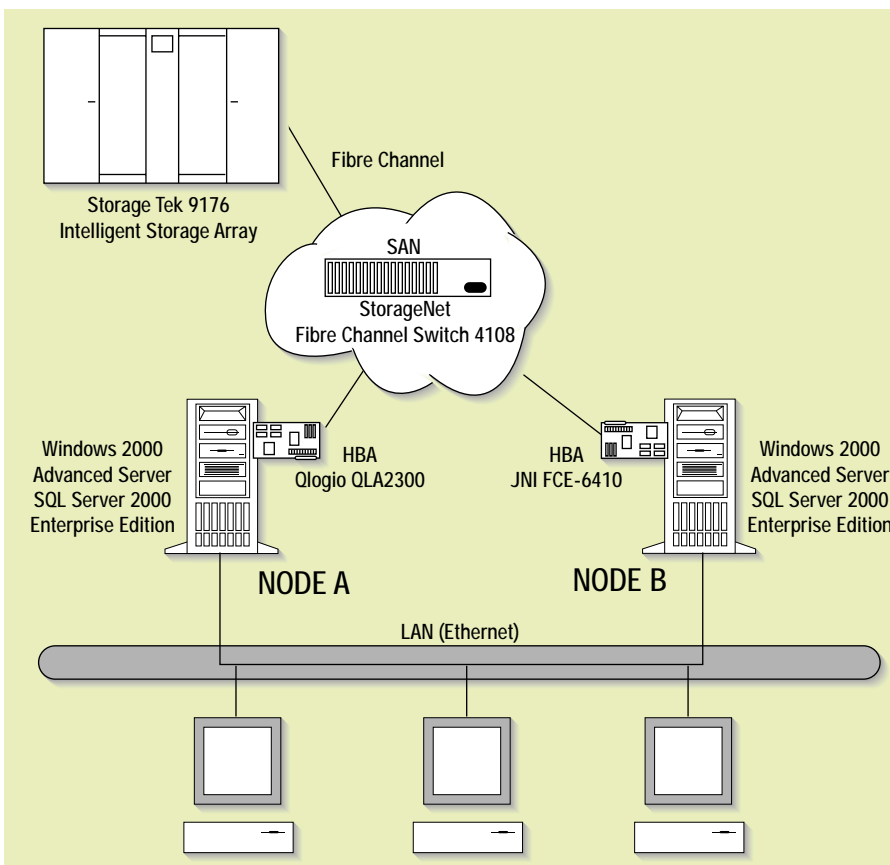
**Onderwerpen van een database aan een benchmark is één ding, analyse van de gemeten waarden is iets anders**

deze methode bruikbaar is om de maximale TPC of BPS te bepalen, is deze niet geschikt voor capaciteitplanning.

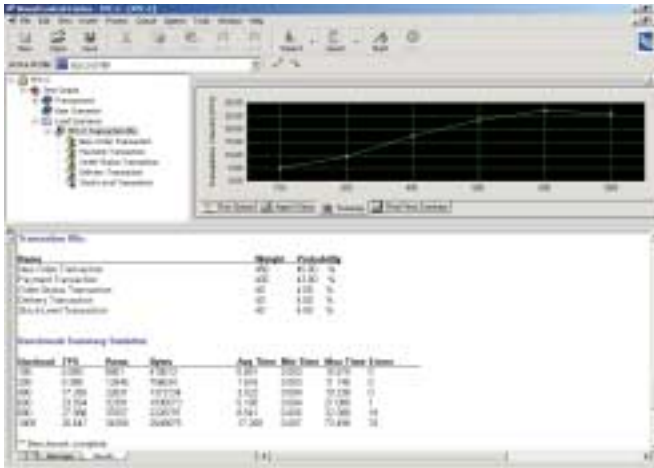
Ten slotte kan de gebruiker zelf twee typen transactions creëren: static en dynamic. Static transactions zijn binnen de benchmarkspecificatie gedefinieerde testscripts en worden vanuit een benchmark-DLL geïmporteerd. De SQL-definities kunnen niet door de gebruiker worden aangepast. Met dynamic transactions kan de gebruiker zijn eigen benchmark transactions creëren.

**TESTOPSTELLING**

Voor onze test maken wij gebruik van SQL Server 2000 Enterprise Edition (met Service Pack 2). We installeren Visual Control Center en een Agent op een Windows 2000 Advanced Server-systeem (Pentium IV 1700 MHz met 512 RAM). SQL Server 2000 draait in een active/passive MSCS-clusteromgeving. Op drie andere Windows 2000-systemen worden agents geïnstalleerd, zodat de database met totaal vier agents kan worden belast. Tijdens de test wordt gebruikersbelasting met honderd gebruikers opgehoogd tot een maximum van duizend. Als testscenario draait de standaard benchmark TPC-C Transaction Mix (zie figuur 5).



**FIGUUR 4: TESTOMGEVING SQL SERVER 2000.**



FIGUUR 5: TPC-C BENCHMARK SQL SERVER 2000.



FIGUUR 6: MET SPOTLIGHT GEDETECTEERDE DEADLOCKS.

Voordat de test kan beginnen, moet de testomgeving worden geconfigureerd. De standaard installatie configureert de Adaptive Server als testdatabase. Om een andere database te testen, moeten verschillende handelingen worden verricht. Ten eerste is het nodig toegang tot de database te krijgen. Met SQL Server kan men daarbij gebruik maken van de -met Windows standaard meegeleverde- ODBC-

**Voor databases zijn verschillende goede benchmarks op de markt, elk met een eigen doel en voor- en nadelen**

drivers. Benchmark Factory brengt alle bij de benchmarktest betrokken instellingen en tests onder in een projectbestand (voor databasebenchmark is dit een .vsb-file). Na aanmaak van dit bestand wordt de Creation Wizard opgestart: een verzameling dialoogboxen die onder meer behulpzaam is bij de aanmaak van de benchmark-objecten. Voor een databasebenchmark betekent dit de creatie van de nodige testtabellen, indexen en stored procedures op de testdatabase. Daarnaast is met deze wizard de 'schaalbaarheid' van de benchmark instelbaar en wordt het eerste contact met de testdatabase gelegd. Schaalbaarheid in de context van database-omgeving duidt op de hoeveelheid gegevens die een database kan bevatten; deze is voor de gebruiker instelbaar - dit om

zoveel mogelijk de werkelijke productie-omgeving te kunnen nabootsen. Het resultaat van de TPC-C benchmark is te zien in het tabblad Summary (zie figuur 5).

Uit de grafiek blijkt dat tot achthonderd gebruikers het aantal transactions per second (tps) maximaal is toegenomen tot ongeveer 27; daarboven de 800 zien we het aantal tps plotseling afnemen; op de mogelijke oorzaak komen we hierna terug. Het aantal tps vertalend naar tpmC, zien we dat het testsysteem slechts 1625 tpmC's haalt.<sup>1</sup> Delen we de prijs van het zelfgebouwde kloonsysteem echter door het aantal tpmC's, dan komen we uit op een Price/tpmC uit die onder de tien dollar ligt, de gunstigste uit de TPC-lijst. Opmerkelijk is dat op de top tien 'TPC-C' lijst met de beste tpmC-waarden de eerste drie plaatsen worden ingenomen door systemen waarop SQL Server Enterprise Edition draait, en op de 'Price/tpmC' lijst bezet het MS-rdbms zelfs alle eerste tien plaatsen.

**FLESSENHALZEN VINDEN**

Een van de doelstellingen van benchmarking is boven water te krijgen welke beperkende factor(en) of flessenhalzen binnen een systeem bestaan. Twee handelingen zijn daarvoor nodig. In de eerste plaats moet het systeem tot zijn *break point* worden belast. Ten tweede is het zaak allerlei systeemfuncties te meten om te bepalen welke component als eerste faalt. Het onderwerpen van een database aan een benchmark is één ding, analyse

van de gemeten waarden is iets anders. Daarin voorziet een benchmarktool niet. Wel kan Benchmark Factory, in combinatie met NT's Performance Monitor, een aantal specifieke statische gegevens verzamelen. Via de bekende Add-knop zijn de extra door Benchmark Factory geïnstalleerde objecten Average Transaction Time, tps en Userload te selecteren.

Naast de extra objecten voor NT's Performance Monitor biedt de firma Quest nog drie speciale tools voor load-testing, capaciteitplanning, bewaking en 'tuning' van databases:

- Spotlight biedt essentiële functies voor tuning van Oracle, SQL Server en web-servers;
- I/Watch, voor het bewaken van essentiële databasefuncties en trendanalyses;
- Foglight - totale bewaking van web- en databaseserverfuncties.

Met Spotlight proberen we te achterhalen wat de reden zou kunnen zijn van de plotselinge terugval in het aantal tps bij 800+ gebruikers. We constateren in het SQL Activity-menu van Spotlight een aanmerkelijke toename van het aantal Deadlocks/sec en Latch Waits/sec. *Lock thrashing* is de meest waarschijnlijke oorzaak. Als het aantal actieve transacties maar hoog genoeg is, heeft iedere volgende gestarte transactie geen kans meer op volledige voltooiing, zonder dat deze door een lock wordt opgehouden. Bij een bepaalde belasting treedt ineens een sneeuwbaaleffect op, waarbij het aantal locks sterk toeneemt. Met als resultaat dat

nog slechts een enkele transactie kan worden beëindigd, wat uitmondt in een sterke teruggang in de databaseprestaties.

Verschillende technieken staan ontwikkelaars en dba ter beschikking om het blocking-gevaar te bezweren. Ook bottlenecks in het systeem zelf kunnen er debet aan zijn, zoals schijf, geheugen of I/O-bandbreedte. Vaak krijgt het systeemgeheugen de schuld en wordt tevergeefs extra geheugen bijgeplaatst. Meestal is een slecht presterend schijfsysteem de oorzaak. Nieuwe, betaalbare technologieën,

zoals *solid state disks* (SSD's) maken het thans mogelijk de toegang tot schijfgeheugen minimaal tienmaal sneller te maken.<sup>2</sup>

### NUT VAN BENCHMARKS

Over het nut van benchmarks zijn de meningen verdeeld. Optimisten zullen zeggen dat benchmarks alles vertellen over de prestaties van een bepaald systeem. Pessimisten verklaren dat benchmarks geen nut hebben, omdat ze de

## Informatie op Internet:

www.quest.com  
www.tpc.org  
www.specbench.org  
www.benchmarkresources.com  
www.etestinglabs.com

'echte' wereld niet kunnen nabootsen. De ene groep zal de rol van benchmarking bagatelliseren, de andere zal verkondigen dat het het enige antwoord is op de vraag hoe een systeem in de praktijk presteert of mogelijk gaat presteren.

Sommige leveranciers van benchmarktools, waaronder Quest Software, geven eerlijk toe dat de waarheid ergens in het midden ligt. Benchmarking is ook geen exacte wetenschap. Helaas staat het nog het meest bij verkoopafdelingen in de belangstelling. Daarmee is benchmarking vaak onterecht in een verkeerd daglicht - dat van 'verdraaiing' en 'interpretatie' van resultaten - komen te staan. Dat maakt benchmarks nog niet waardeloos. Je moet ze nemen voor wat ze zijn. Ze bieden een waardevolle informatiebron en geven een reëel inzicht in de staat van de IT-technologie als geheel. ●

### Noten en literatuur:

1. Dit steekt natuurlijk schril af tegen de nummer één van de officiële TPC-lijst: de Compaq Proliant met 709,220 tpmC. Zie <http://www.tpc.org/tpcc/default.asp>
2. Zie het thema-artikel over dit onderwerp in het vorige nummer van DB/M.

Artikelen in *SQL Server Magazine*:

*How Valuable Are Benchmark Scores*, [www.sqlmag.com/Articles/Print.cfm?ArticleID=24467](http://www.sqlmag.com/Articles/Print.cfm?ArticleID=24467)

*Benchmark Results and Ketchup*, [www.sqlmag.com/Articles/Index.cfm?ArticleID=20230](http://www.sqlmag.com/Articles/Index.cfm?ArticleID=20230)

*Windows, SQL Server Set Another TPC-C record*, [www.sqlmag.com/Articles/Index.cfm?ArticleID=23228](http://www.sqlmag.com/Articles/Index.cfm?ArticleID=23228)

Bram Dons (b.dons@chello.nl) is onafhankelijk IT-adviseur.

## Databasebenchmarks

Afhankelijk van de omgeving waarin ze worden toegepast, kunnen benchmarks enorm van elkaar verschillen, bijvoorbeeld voor zakelijke of wetenschappelijke applicaties.

Toch worden goede benchmarks gekenmerkt door de volgende eigenschappen:

- simulatie van veel gebruikers met minimale hardware;
- tonen van resultaat in verschillende vormen;
- herhaalbaar - opnieuw draaien van de test moet hetzelfde resultaat opleveren;
- schaalbaar, leverancierneutraal;
- rekening houdend met veel variabelen.

Voor databases zijn verschillende goede benchmarks op de markt; elk met een eigen doel en voor- en nadelen. De meest bekende zijn:

- TPC-B, C en D. De uit 1988 stammende TPC had tot doel een meetstandaard voor de OLTP-omgeving te ontwikkelen. De verouderde TPC-B test meet de doorvoer in transactions per second (tps) en is ontworpen voor tests van de basiscomponenten van een databasesysteem. De verbeterde TPC-C simuleert een meer gecompliceerde omgeving met 'echte' applicaties. In TPC-C is 'throughput' gedefinieerd als de hoeveelheid nieuwe 'New-Order' transactions per minute (tpmC) die een systeem genereert, terwijl het op hetzelfde tijdstip vier andere typen transacties uitvoert (Payment, Order-Status, Delivery en Stock-Level). TPC-D werd in april 1995 door het TPC aanvaard. Het verschilt van de vorige versies, die meer gericht zijn op lichte en middelgrote klantgerichte transacties. TPC-D bevat een grote variëteit aan decision-support-applicatietaken die complexe, langdurende query's draaien in combinatie met complexe datastructuren;
- AS3AP (ANSI SQL Standard Scalable and Portable), ontwikkeld op de Cornell Universiteit in samenwerking met Carolyn Turbyfill en Cyril Orji. Het is ontwikkeld als algemene benchmark voor de vergelijking van relationele databases op verschillende architecturen. AS3AP-tests worden in twee modules opgedeeld. Er is een *single-user* test, voor het laden en structuren van de database en het genereren van de query's om de toegangsmethodes en basis-query-optimalisatie te testen. De *multi-user* test dient voor OLTP-werkbelasting, opvragen van informatie en gemengde werkbelasting;
- Ziff-Davis Database en Ritesize Benchmark. Ziff-Davis koos een subset van de AS3AP-benchmarktests om rdbms'en te testen. Deze tests bevatten een Load en Index-, Random Write-, Ad Hoc Query- en Mixed Workload-test. De benchmark is verder verbeterd door de mogelijkheid grafische bestanden op te slaan, inclusief BLOB-gegevenstypen. De grote database-vendors nemen deze benchmarktests over algemeen wat minder serieus.