

Alle eigenschappen voor continue toegang tot data zijn aanwezig

SQL Server 2008 Release Candidate 0

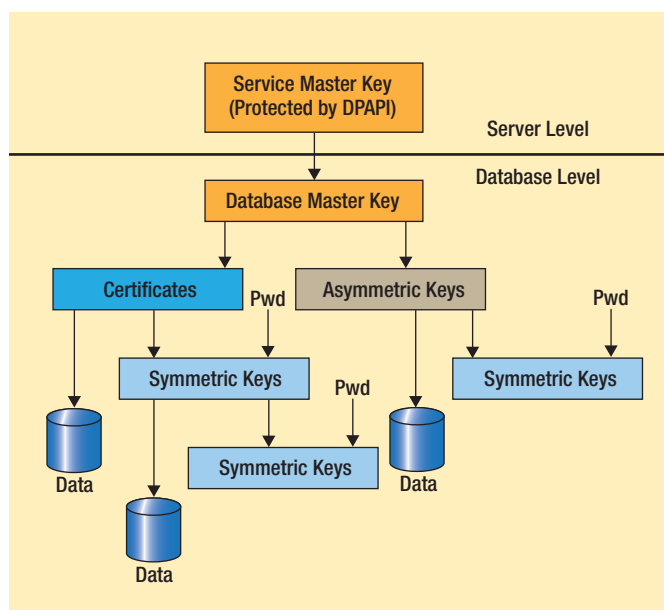
Bram Dons

Sinds kort is de nieuwe SQL Server 2008 als RC0 beschikbaar waarvan een trial-versie is te downloaden van de Microsoft website. SQL Server 2008 moet volgens Microsoft een Data Platform bieden dat elke soort data, op elke plaats, en op elk moment kan beheren. Het is in staat om gestructureerde, semi- en ongestructureerde data, zoals images en muziek, direct in een relationele database op te slaan.

SQL Server 2008 ondersteunt een rijke verzameling aan geïntegreerde services zoals opvragen, zoeken, synchroniseren, rapportage en analyse van data. De database kan overweg met zakelijke applicaties die zijn ontwikkeld op basis van Microsoft .NET, Visual Studio en binnen een service georiënteerde architectuur (SOA) en business process via Microsoft BizTalk Server. SQL Server 2008 biedt op verschillende terreinen talrijke nieuwe features, waaronder databeveiliging, business continuity, voorspelbare respons, prestaties, schaalbaarheid, beheer, applicatie-ontwikkeling, data-opslag en -integratie, datawarehousing, service analyse, datamining en rapportage. De belangrijkste features passeren de revue.

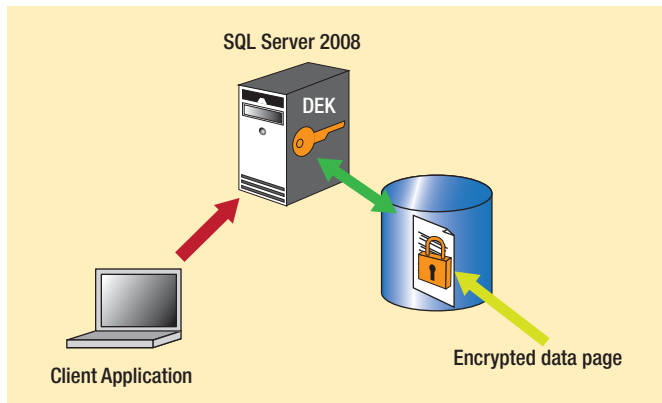
Beveiliging

SQL Server 2008 borduurt voort op het belangrijke principe van *least privilege*, waarmee gebruikers niet meer rechten worden gegeven dan strikt noodzakelijk is om hun werk goed te kunnen uitvoeren. Er is al veel geschreven over Microsoft's Thrustworthy Computing, waarvan de belangrijkste vier pijlers zijn: secure by design, default, deployment en communications. Kernelement van Thrustworthy Computing is het *secure by default* mandaat, wat betekent dat SQL Server 2008 bij de installatie al veilig 'out of the box' is met veilige default-instellingen. Zoals gezegd, deze richtlijnen zijn op verschillende terreinen ook toegepast in SQL Server 2008. Al bij installatie controleert de Configuration Checker op eventuele tekortkomingen in de beveiliging. De meeste beveiligingsfeatures zijn dan nog niet geactiveerd of geïnstalleerd, zoals CLR-integratie, database mirroring, debugging en Service Broker. Een van de nieuwigheden is het *Surface Area* facet waarmee een policy is te definiëren die de status van verschillende SQL Server 2008 features beheert. Daarmee is de gewenste 'surface area' in te stellen voor servers waarmee een minimale surface area op alle SQL Server instances binnen de onderneming kan worden afgedwongen, wat het risico op inbraak moet reduceren.



Afbeelding 1: Encryptie-key hiërarchie in SQL Server 2008.

Verder is bij SQL Server 2008 de SQL Server Authenticatie-optie verbeterd. Ten eerste ondersteunt het default de verbinding door middel van SQL-gegenereerde certificaten. Database-beheerders hoeven geen valide SSL certificaat aan te vragen en te installeren, omdat SQL Server 2008 deze automatisch genereert en het kanaal automatisch versleutelt. Met SQL Server 2008 is password policy enforcement in de server ingebouwd. De server ondersteunt de traditionele, binaire Tabular Data Stream voor client-toegang en native XML Web service-toegang via http. Het belangrijkste voordeel bij toegang via http is dat iedere client-software en ontwikkel-tool, die overweg kan met webservice-protocollen, toegang heeft tot data op de SQL Server. Dit betekent dat SQL Server 2008 kan voorzien in stand-alone webservice-methoden alsmede een complete endpoint in een Service Oriented Architecture. SQL Server 2008 ondersteunt vijf verschillende authenticatietypen: Basic, Digest, NTLM, Kerberos en Integrated authenticatie en heeft uitvoerige mogelijkheden om data te versleutelen via symmetrische- en asymmetrische sleutels en digitale certificaten.



Afbeelding 2: Transparent Data Encryption (bron Microsoft).

In SQL Server 2005 was het al mogelijk om data in de database te versleutelen door een custom Transact-SQL te schrijven. SQL Server 2008 maakt dit proces eenvoudiger door de introductie van Transparent Data Encryption. Deze voert alle kryptografische bewerkingen op database-niveau uit zodat de applicatie-ontwikkelaar geen enkele code meer hoeft te schrijven. Op het moment dat de data naar disk worden geschreven, vindt de versleuteling plaats en omgekeerd bij het lezen de ontsleuteling. De Transparent Data Encryption methode biedt als voordeel dat een database is te beveiligen zonder dat bestaande applicaties hoeven te worden aangepast. Bij de versleuteling wordt van een Database Encryption Key (DEK) gebruik gemaakt, waarbij de DEK in de database boot record wordt opgeslagen zodat deze beschikbaar blijft ingeval van een recovery van de database. Men kan van een Service Master Key of Hardware Security Module (HSM) gebruik maken om de DEK te beschermen; HSM's zijn meestal USB-devices of smartcards.

Datawarehouse schaalbaarheid

Nummer één op het ontwikkelingslijstje van de SQL Server 2008 release stond de verbetering van de schaalbaarheid van de complete SQL Server product suite, om daarmee tegemoet te komen aan de toenemende eisen van grote ondernemingen. Belangrijke verbeteringen op datagebied die aan SQL Server 2008 zijn aangebracht betreffen niet alleen de database engine zelf, maar ook verbeteringen op het gebied van integration-, analysis en reporting services. Aangebrachte verbeteringen op het relationele DBMS datawarehouse zijn onder meer Star Join, Partitioned Table Parallelism (PTP), Data Compression en Resource Governor. Bij dimensioneel gemodelleerde datawarehouses bestaat een groot deel van de werkbelasting uit wat bekend staat als star join query's. Dergelijke query's volgen een algemeen patroon dat de fact tabel met een of meerdere dimension tabellen samenvoegt. SQL Server 2008 biedt snelheidsverbeteringen voor veel star join query's die een belangrijke deel van de fact tabelrijen verwerken. De nieuwe technologie is gebaseerd op bitmap filters, die ook wel bekend staan als *Bloom filters*. Daarmee is de SQL Server in staat om niet-gekwalficeerde fact tabelrijen tijdens een vroegtijdig stadium van het query-evaluatieproces te

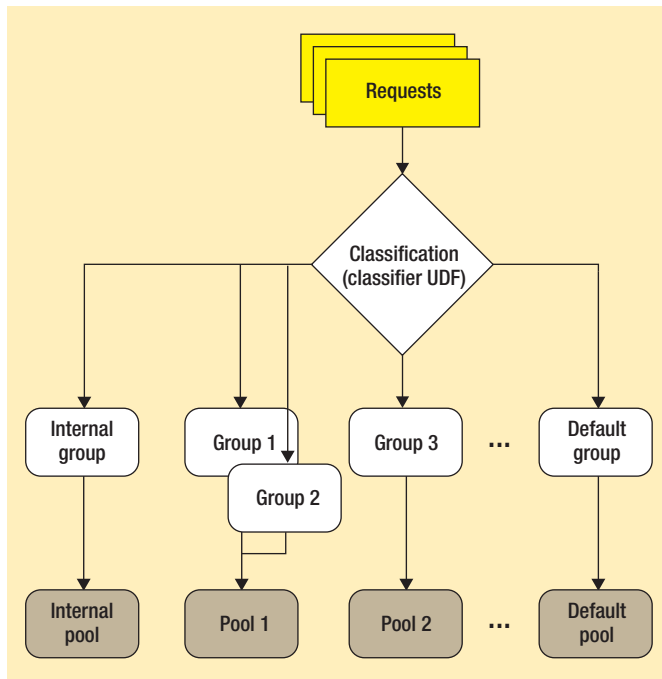
eliminieren. Dit bespaart een aanzienlijke hoeveelheid CPU-tijd, in vergelijking met query processing-technologieën die door de concurrentie worden gebruikt. Microsoft claimt in een van haar white papers een verbetering van 15 tot 25 procent en bij sommige query's is zelfs een factor zeven of meer aan verbetering geconstateerd.

In SQL Server 2005 gebruiken query's die meer dan een partitie benaderen een thread, wat gelijk staat aan een processor core. Dit vormt soms een beperking voor de query prestaties, vooral bij gebruik in Shared Memory Multiprocessor (SMP) computers met veel processor cores. De Partitioned Table Parallelism (PTP) feature van SQL Server 2008 verbetert de prestaties door beter gebruik te maken van de beschikbare hardware, ongeacht hoeveel partities er worden benaderd. De PTP feature is default ingeschakeld, zonder dat de beheerder deze handmatig hoeft te configureren of tunen. Een andere feature met betrekking tot partities is de toepassing van partition-aligned geïndexeerde views. Dat biedt de voordelen van een geïndexeerde view bij grote gepartitioneerde tabellen, waar voorheen het bezwaar gold van het opnieuw moeten samenstellen van een complete samengestelde gepartitioneerde tabel. Met partition-aligned views kan dat nu worden voorkomen.

Change Data Capture (CDC) is een nieuwe data tracking feature van SQL Server 2008 dat hoofdzakelijk is ontworpen als een efficiënt systeem om veranderde data in gebruikerstabellen te kunnen volgen. Het biedt de gebruiker toegang tot deze data in een gemakkelijk 'consumeerbaar' relationeel format. Zo is CDC bruikbaar in een operationele database om veranderde data bij te houden die naderhand naar een datawarehouse kunnen worden verplaatst. De beschikbaarheid van CDC in SQL Server neemt de noodzaak weg om belastende methoden als gebruikers-triggers, timestamp columns en 'dure' query's uit te moeten voeren, om vast te stellen welke data in de operationele database zijn veranderd.

SQL Server 2008 kan voorzien in stand-alone webservice-methoden

De nieuwe datacompressie feature reduceert de grootte van tabellen, indexen of subsets van hun partities, door datatypes met een vaste lengte in variabele lengte format op te slaan en de reductie van redundante data. De bereikte ruimtebesparing is wel afhankelijk van het schema en de mate van datadistributie. Microsoft claimt bij het testen van verschillende datawarehouse databases een reductie van echte gebruikerstabellen tot 87 procent (een 7-op-1 compressieverhouding) te hebben waargeno-



Afbeelding 3: Voorbeeld Resource Governor (bron Microsoft).

men. Maar in het algemeen kan een reductie in het 50 tot 70 procent (compressieverhouding 2-op-1 en 3-op-1) bereik worden verwacht. SQL Server 2008 ondersteunt twee typen compressie. ROW compressie slaat data types met vaste lengte op in variabelen en het andere type, PAGE compressie, is gebouwd op basis van ROW compressie en slaat veel voorkomende byte-patternen één keer op met verwijzingen naar waarden in de respectievelijke kolommen. Maar datacompressie vraagt zijn tol: een verhoging van de CPU-belasting. Deze overhead wordt merkbaar bij een query of de uitvoering van een DML bewerking op gecomprimeerde data. De relatieve CPU-overhead met ROW is minder dan voor PAGE, maar PAGE compressie biedt een betere compressie. De database-beheerder kan kiezen om de hele tabel te comprimeren of een subset van een partitie. Afhankelijk van de CPU- en I/O-'kosten' kan in een datawarehouse-omgeving worden gekozen tussen ROW en PAGE.

Resource Governor

De nieuwe Resource Governor in SQL Server 2008 maakt het mogelijk om de hoeveelheid CPU en geheugenbronnen te alloceren aan verschillende delen van de relationele database werkbelasting. Een toepassing is het voorkomen van 'runaway' query's (waarbij andere gebruikers de toegang tot deze bronnen wordt ontzegd) en om computerbronnen te reserveren voor belangrijke delen van de werkbelasting. In SQL Server 2005 werden alle werkbelastingen door allocatie policy's hetzelfde behandeld en alle gedeelde bronnen (bijvoorbeeld, CPU-bandbreedte en geheugen) op het moment dat er een request werd uitgevaardigd. Dit leidde soms tot een disproportionele distributie van computerbronnen, wat op zich weer resulteerde in

een onevenwichtige verhouding tussen de werkbelasting en bronnen of onverwachte vertragingen. Er zijn drie nieuwe concepten geïntroduceerd bij de Resource Governor. Ten eerste de GROUP, een workload group of group, en is een door de gebruiker gedefinieerde categorie met requests. Ten tweede een POOL, resource pool of pool, die een deel van de fysieke bronnen op de server vertegenwoordigt. De pool kan een vaste grootte hebben of deel uit maken van een share tussen meerdere pools. Tenslotte Classification, dat is een verzameling door de gebruiker geschreven regels die de Resource Governor in staat stellen om de requests te classificeren in de hiervoor beschreven groepen.

Lookup prestaties

Het verplaatsen van informatie tussen het operationele systeem naar een datawarehouse kan een tijdsintensieve taak zijn. Om dit proces te versnellen zijn er met SQL Server 2008 Integration Services (SSIS) twee belangrijke schaalbaarheids-features geïntroduceerd: verbeterde lookup en transformation pipelineprestaties. De Lookup component in SSIS draait snel en is eenvoudiger te programmeren dan in SQL Server 2005 het geval was. Een Lookup operatie bouwt een cache op met opgevraagde rijen uit een bepaalde dataset. In SQL Server 2005 waren data alleen van een specifieke OleDb verbinding op te vragen en de cache alleen te vullen door middel van een SQL query. Bij SQL Server 2008 gaat het vullen via een aparte pipeline. SQL Server 2005 herlaadt de cache ieder keer wanneer het wordt gebruikt, bij SQL Server 2008 kan de cache in virtueel geheugen of permanent in file storage worden opslagen, zodat meerdere Lookup componenten dezelfde cache-inhoud kunnen delen. Een andere nieuwe feature van Lookup is de miss-cache. Wanneer er een Lookup actie direct op de database wordt uitgevoerd dan bespaart de miss-cache feature tijd door optioneel de sleutelwaarden die geen gelijke waarden in de gerefereerde dataset bevatten in cache op te slaan.

De nieuwe datacompressie feature reduceert de grootte van tabellen

Verder kunnen in SQL Server 2008 SSIS meerdere threads met elkaar samenwerken om het werk te doen dat, zoals in SQL Server 2005, noodgedwongen door een enkele thread gedaan moesten worden. Om een hoog niveau van parallelisme te bereiken biedt SSIS meer mogelijkheden tot parallelverwerking door gebruik van multiprocessor-systemen.

Schaalbare Shared Database voor AS

SQL Server 2008 Analysis Services ondersteunt met een nieuwe feature, Scalable Shared Database (SSD) genaamd, de schaal-

baarheid van OLAP query werkbelasting bij toepassing van meerdere kleine servers en met slechts een enkele kopie van de database. SSD is in verschillende scenario's toepasbaar, maar de meest gebruikelijke is een scale-out scenario bij toepassing van Analysis Services om bij een toenemend aantal gelijktijdige gebruikers de query-belasting over meerdere servers te verdelen. In het verleden werd dit bereikt door middel van een load balancing oplossing als Microsoft Network Load Balancing (NLB). Maar deze bevond zich aan de voorkant van de servers en repliceerde de fysieke data tussen de servers, met als nadeel dat een dergelijke omgeving op talrijke gebieden lastig is te beheren (waarvan datareplicatie er een is). SSD biedt database-beheerders de mogelijkheid om een database als 'read-only' te markeren en via een Storage Area Network (SAN) door meerdere server instances te laten delen. Dit bespaart diskruimte en de tijd die anders nodig zou zijn voor datareplicatie. Een alternatieve oplossing om de prestaties te verbeteren is scale-up, waarbij een enkele grote server wordt gebruikt, in plaats van talrijke kleine. Voordeel van scale-up is dat een enkele query sneller kan worden uitgevoerd op een groot systeem.

Voordeel van scale-up is dat een enkele query sneller kan worden uitgevoerd op een groot systeem

Maar scale-out met SSD kan de kosten van hardware reduceren (gegeven de lagere kosten per processor) en nog steeds voldoen aan de vraag bij veel gelijktijdige gebruikerswerkbelastingen. Verder is de schaalbaarheid van processoren met SSD veel hoger dan bij toepassing van een enkele grote server.

Scale-up met SQL Server 2008

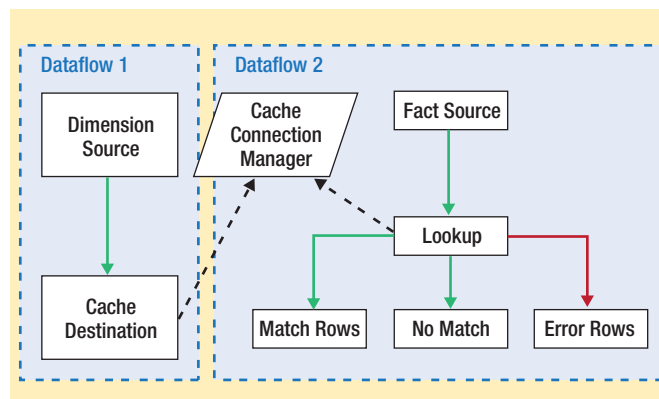
Server-consolidatie, grote data-opslagsystemen en complexe query's vereisen fysieke bronnen die de verschillende belastingen kunnen ondersteunen op een server. SQL Server 2008 is in staat om van de laatste hardware-technologieën te kunnen profiteren. Voor hardware-consolidatiedoelinden kunnen meerdere database engine instances en meerdere analyse-services op een enkele server draaien. Niet minder dan 50 instances kunnen op een enkele server worden geïnstalleerd zonder dat hoeft te worden ingeboet op prestaties en responstijden. SQL Server 2008 kan volledig gebruik maken van de laatste CPU-technologie, waaronder 64-bit, multi-core, en multi-processor systemen. SQL Server kan 64 GB geheugen adresseren en ondersteunt de dynamische toewijzing van AWE-mapped geheugen op 32-bit hardware en kan tot 8 TB geheugen op 64-bit hardware gebruiken. Wanneer grote aantallen processoren aan een server worden toegevoegd, dan kan de toegang tot het geheugen vertragen, als de processoren geheugen moeten adresseren dat

niet lokaal aan de processor is verbonden. Hardware op basis van Non-Uniform Memory Access (NUMA) architectuur heft dit indirecte geheugen toegangsprobleem op door alle aanwezige processoren toegang te geven tot lokaal geheugen. SQL Server 2008 is 'NUMA-aware' waardoor het kan profiteren van een grotere schaalbaarheid, waarbij NUMA-gebaseerde computers zonder veranderingen aan de applicatie kunnen worden toegepast. SQL Server 2008 ondersteunt zowel hardware- als soft-NUMA. SQL Server 2008 ondersteunt soft-NUMA voor servers die wel veel CPU's bevatten maar geen hardware NUMA ondersteunen; met soft-NUMA wordt alleen de SQL Server Network Interface (SNI) en scheduler ondersteund.

Alhoewel de schaalbaarheid van een SQL Server instance eenvoudig is te vergroten door meer geheugen of CPU's toe te voegen, moet daarvoor wel de server worden uitgeschakeld, wat in een 24/7-omgeving niet tolerabel is. Met SQL Server 2008 zijn CPU's en geheugen bij compatibele servers toe te voegen, zonder dat de database services hoeven te worden gestopt. Voor de toevoeging van hot-add memory worden de volgende eisen gesteld: SQL Server 2008 Enterprise Edition; Windows Server 2002 Enterprise of Datacenter Edition; hardware dat toevoegen van geheugen of virtualisatie-software ondersteunt en SQL Server gestart met de -h optie. Voor hot-add CPU's gelden de volgende operating system-eisen: SQL Server 2008 Enterprise; Windows Server 2008 Enterprise Edition voor Itanium Systems of Windows Server 2008 Datacenter Edition voor x64 bit systemen; 64-bit SQL Server; hardware CPU toevoeging of virtualisatie software.

SQL Server 2008 High Availability

Hiervoor zagen we welke belangrijke verbeteringen zijn aangebracht op het gebied van beveiliging en schaalbaarheid. Maar minstens zo belangrijk in een high-end database omgeving is beschikbaarheid. Want steeds meer database-applicaties worden 'mission-critical'. Dat houdt in dat ze een minimale downtime vragen en een snel herstel, ingeval van een catastrofe. SQL Server biedt met 'Always On Technologies' een scala aan voorzieningen om de downtime te minimaliseren en een hoge

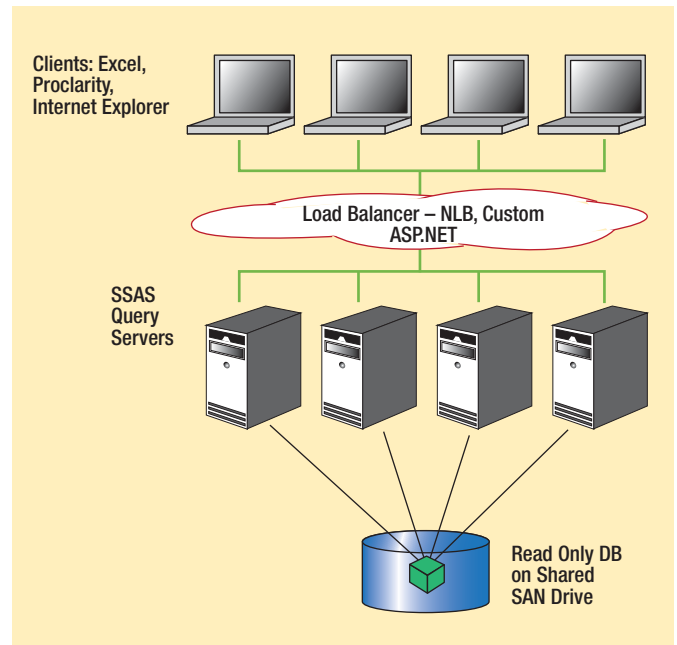


Afbeelding 4: Lookup scenario (bron Microsoft).

mate van beschikbaarheid te bieden die we in het kort zullen benoemen (in een vervolgartikel gaan we uitvoerig in op de Always On Technologie en de praktische implementatie daarvan). De in SQL Server 2005 ondersteunde Database Mirroring is in SQL Server 2008 uitgebreid met compressie en men hoeft een database niet langer te herstarten in geval van een handmatige failover. Ook de noodzaak komt te vervallen om een drive letter voor elke SQL Server instance te moeten toewijzen, wat het aantal mogelijke instances bij een geclusterde oplossing verhoogt. SQL Server 2008 introduceert verbetering aan de peer-to-peer replicatie, door nodes te kunnen toevoegen terwijl het replicatieproces online is. Door SQL Server 2008 te combineren met Windows Server 2008 wordt een solide HA-oplossing voor een totale SQL Server instance geboden.

Epiloog

Met de uiteindelijke release van SQL Server 2008 denkt Microsoft een database te kunnen bieden die geschikt is voor het datawarehouse, transaction processing en business-critical applicaties in een high-end data center. Op het gebied van beveiliging, beschikbaarheid, consolidatie en prestaties zijn daartoe de nodige aanpassingen en verbeteringen aangebracht. Naast deze, voor een datacenter database, belangrijke eisen, zijn er ook op het gebied van programmering, beheer, datatypen, analyse, datapresentatie en query allerlei nieuwe features geïmplementeerd. Het door Microsoft genoemde 'Data Platform' komt daarmee aan de vraag tegemoet om de toenemende verscheidenheid aan datatypen in een relationele database te kunnen opslaan, waaronder XML, e-mail, tijd/datum, geospatial, enzovoort. Met de komst van SQL Server 2008 denkt Microsoft



Afbeelding 5: Schaalbare, gedeelde database voor AS (bron Microsoft).

een database te kunnen bieden die alle eigenschappen heeft die een onderneming vraagt voor het bieden van een continue toegang tot data, waarbij beveiliging, betrouwbaarheid en schaalbaarheid voor mission critical applicaties hoog in het vaandel staan.

SQL Server 2008 RTM is te downloaden vanaf:

www.microsoft.com/sqlserver/2008/en/us/trial-software.aspx

Bram Dons is onafhankelijk IT consultant.

Thema Business Rules

Vervolg van pagina 31

Beter voldoen

Bedrijfsregels in de vorm van complexe volzinnen worden omgezet in kleinere samenhangende kennisregels en beperkingregels. Elk van deze regels kan makkelijk uitgelegd worden aan de eindgebruikers. Zo ook de samenhang tussen deze regels. En op welke informatie de regels inspelen en wanneer. Er wordt dus voldaan aan drie belangrijke aspecten die de Business Rules Approach oplegt, namelijk het begrijpelijk maken van de bedrijfsvoering voor de eindgebruiker, de traceerbaarheid van waar en wanneer bedrijfsregels uitgevoerd en/of gecontroleerd worden, en de scheiding van gegevens, processen en bedrijfsregels. Bovendien zijn deze kleine samenhangende regels beter onderhoudbaar dan grote complexe volzinnen, en komt men gemakkelijker tot hergebruik van regels.

Als we de aanhangers mogen geloven, zal deze aanpak zorgen voor systemen die beter voldoen aan de noden van de eindge-

bruikers. En het creëren van applicaties zou minder tijd in beslag nemen dan nu het geval is. Om maar te zwijgen over het onderhoud aan systemen. Daar zou een winst van meer dan 50 procent geen utopie zijn. En dat is geen overbodige luxe bij de huidige dynamiek die bedrijven en organisaties nastreven bij hun bedrijfsvoering.

Noten

1. Barbara von Halle, *Business Rules Applied, Building Better Systems Using the Business Rules Approach.*
2. R.J. Veldwijk, *10 Geboden voor Goed Database-ontwerp.*
3. C.J. Date, *What Not How: The Business Rules Approach to Application Development.*

Marc Dierick (marc@mardinfo.be) is onafhankelijk IT-consultant bij Mardinfo.