

Alterian, Omnidex, RapidBase en TimesTen weinig bekende producten

# Zoektocht naar onbekende databases

Robbert Hoeffnagel

**Commercieel mogen ze dan weinig tot niets voorstellen, er bestaan wel degelijk uiterst interessante alternatieven voor Oracle, SQL Server of DB2. Alterian bijvoorbeeld. En wat te denken van Omnidex, RapidBase of TimesTen? De namen zullen vaak wel enigszins bekend voorkomen. Maar wat was er ook alweer speciaal aan deze producten? Sterker nog: bestaan ze eigenlijk nog wel? Database Magazine ging op onderzoek uit – met wisselend resultaat.**

Al bij de eerste zoekopdracht die bij Google wordt ingetikt, gaat het fout. 'RapidBase' lijkt een flinke lijst met hits op te leveren. Bovenaan staat zowaar een link waarboven pontificaal staat: rapidbase. Juich echter niet te snel: de pagina waar deze link naartoe leidt bestaat niet. De foutmelding staat op een pagina die behoort tot een domein van een organisatie: VTT, een Fins onderzoeksinstituut.

## Industriële toepassingen

Via de zoekfunctie op deze site komen we het document 'Design of RapidBase – an active measurement system' op het spoor. Het blijkt een conferentie-paper te zijn waarvoor echter betaald moet worden. De bijbehorende 'abstract' levert daarentegen wel wat meer informatie op:

"In data-intensive industrial on-line applications utilizing live process data, one faces an unusual set of database requirements. The process measurement data need to be acquired at great speed, organized in time series and made available for time-based retrieval. Active capabilities and functional extensibility are needed to implement a flexible data-driven processing paradigm. An efficient transaction logging and recovery mechanism is needed in order not to impede the data acquisition flow. RapidBase is a system that meets these requirements."

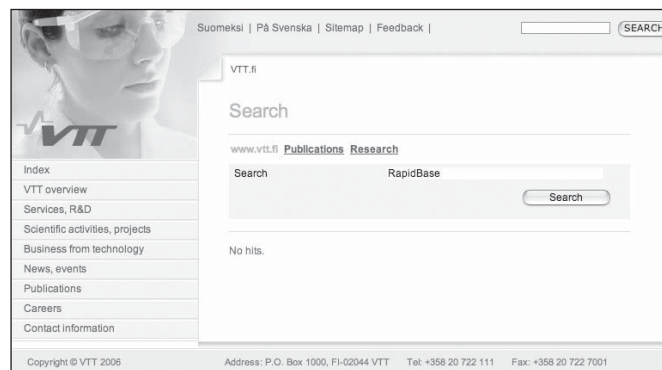
## Geen reactie

Een van de auteurs van de paper is ene Antoni Wolski, een man die ook een van de auteurs is van twee papers getiteld 'Performance measurement and tuning of hot-standby databases' en 'Quantified and temporal fuzzy reasoning for active monitoring in RapidBase'. In beide gevallen gaat het om een wetenschappelijke verhandeling, waarbij wiskundige formules niet geschuwd worden en deze handelen over het gebruik van database-technologie in industriële processen.

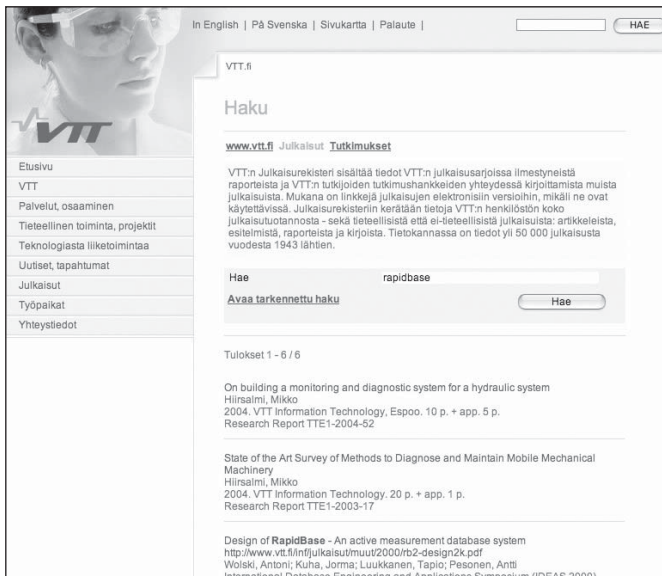
De meeste informatie die online over RapidBase te vinden is, komt uit de periode rond het jaar 2000. Helaas lijkt er echter geen web-pagina te bestaan waar simpelweg RapidBase als software-programma wordt beschreven of kan worden gedownload. Weliswaar vindt Google een pagina waar dit volgens de korte tekst bij de link wel mogelijk zou moeten zijn, maar die pagina blijkt vervolgens bij VTT niet meer online te staan. Mailtjes naar de op VTT-site aangegeven contactpersoon – genaamd Kirjaamo – waren bij het ter perse gaan van dit nummer van Database Magazine nog altijd niet beantwoord.

## TimesTen

TimesTen laat zich daarentegen wel uiterst gemakkelijk vinden. Oracle heeft er zelfs een aparte website aan gewijd. Sinds 20 juni 2005 is dit concern eigenaar van TimesTen. Het gaat hier om software voor real-time datamanagement. Oracle gebruikt de technologie van TimesTen inmiddels als basis voor wat tegenwoordig de 'Oracle TimesTen in-memory database' heet. Er wordt hierbij een architectuur toegepast die geheel geoptimaliseerd is voor gebruik van de database direct in het werkgeheugen van een



Afbeelding 1: Op zoek naar RapidBase.



**Afbeelding 2:** Lijst met op de VTT-site gevonden documenten waarin het trefwoord 'RapidBase' voorkomt.

server. Het voordeel hiervan is uiteraard dat de snelheid van werken vele malen hoger is dan wanneer sprake is van een database die grotendeels op disk staat.

## Over cache en VOIP

Net als bij RapidBase, hebben we hier te maken met toepassingen waarbij het vastleggen van of toegang bieden tot data zeer snel dient plaats te vinden. Het gaat dan om toepassingen in industriële processen, de telecommunicatie en dergelijke. De nieuwe eigenaar ziet in TimesTen echter ook een soort cache voor Oracle-data. Daar is in de vorm van Cache Connect to Oracle zelfs een apart product voor ontwikkeld. Dit is een real-time cache die men zonder problemen kan updaten, zodat data die in reguliere Oracle-databases zijn ondergebracht toch in real-time of in ieder geval zeer responsieve applicaties kunnen worden toegepast.

## Interessant is ook het fenomeen 'instant counts'

De website die Oracle voor dit product heeft opgezet, is zeer informatief. Interessant is onder andere een aantal PDF's en webcasts, waarin ingegaan wordt op de rol die deze in-memory database kan spelen voor *telco's* en service providers die toepassingen als 'voice over ip' (VOIP) of 'location based services' aan klanten willen aanbieden.

TimesTen is voor Oracle een niche-product, maar wel eentje in een erg interessant marktsegment. Want als ergens de grenzen van de (database)technologie worden opgezocht, dan is het wel in de wereld van de real-time enterprise. De database kan gratis worden gedownload en is beschikbaar voor een handvol platformen. Maar wellicht is het toch verstandiger om eerst nog even een kijkje te

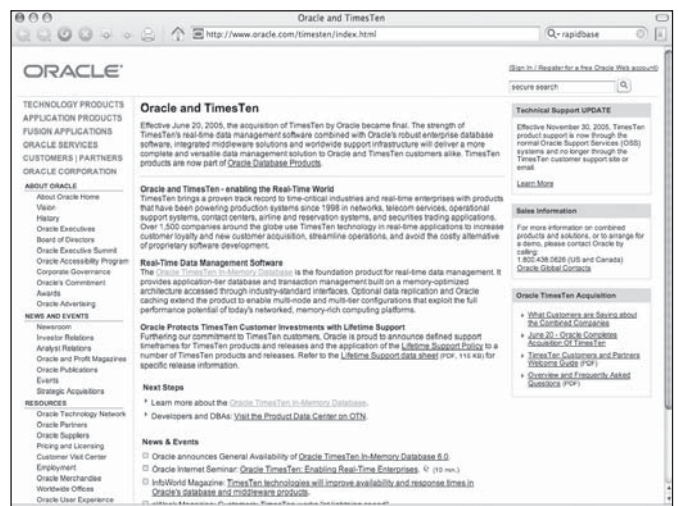
nemen bij een vergelijking die op deze site te vinden is en die TimesTen tegenover BerkeleyDB en Oracle Lite Edition plaatst.

## Omnidex

Google schotelt na het intikken van 'Omnidex' direct de website van DISC voor. Maar dat Dynamic Information Systems Corp de drijvende kracht achter 'the corporate database search engine' is, wisten we eerlijk gezegd al. DISC is een klein bedrijf met wereldwijd slechts twee vestigingen: het hoofdkantoor in Boulder, Colorado en een Europees kantoor in het Britse Leatherhead. Het bedrijf bestaat dit jaar precies een kwart eeuw en houdt zich bezig met het ontwikkelen van wat het noemt 'database indexing software'. Omnidex wordt gepositioneerd als een uiterst schaalbare zoekmachine voor databases. Op basis van een combinatie van in eigen beheer ontwikkelde indexeringstechnieken en op tekst gebaseerde zoekmethoden kunnen grote gegevensverzamelingen worden ontsloten. Hoe die verzamelingen zijn georganiseerd, is eigenlijk niet zo interessant. Het mogen tabellen in relationele databases zijn, maar ook rechttoe rechtaan bestandssystemen, ASCII-, HTML- of XML-documenten.

## 500 miljoen

De belofte van Omnidex is een zeer snelle en bovendien accurate toegang tot data. Daarbij kunnen bestaande databases worden gebruikt. Het bedrijf legt hier een laag indexeringssoftware overheen die – wat het bedrijf enigszins ronkend noemt – 'multi-dimensional search' mogelijk maakt. Hoe dat precies werkt, laat men min of meer in het midden. Wel wordt duidelijk dat er een combinatie wordt gezocht van snelle toegang tot data via de toegepaste indexeringsmethode, terwijl tegelijkertijd – en dat klinkt natuurlijk interessant – 'decision support systems' worden gebruikt. Een derde component luistert naar de naam 'optimized precision text search'. Maar ook die kreet wordt helaas niet veel verder uitgewerkt. Wel legt het bedrijf een toch stevig te noemen snelheidsclaim neer: tot 500 miljoen records per seconde. Nu is het niet waar dat DISC helemaal niets verteld over de



**Afbeelding 3:** Oracle's informatieve site over TimesTen.

Feature	TimesTen	Berkeley DB	Lite Edition
<b>Linked into the application</b> Run in the same process as the application and store data locally instead of on a remote server	✓	✓	✓
<b>Zero administration</b> Allow programmatic administration instead of requiring human administration	✓	✓	✓
<b>Very high performance</b> Optimized and tunable for very high performance. Runs in-memory.	✓	✓	
<b>Real-time replication</b> Replicate databases for higher performance, scalability and availability	✓	✓	
<b>SQL support</b> Supports SQL-92	✓		✓
<b>XQuery support</b> Support XQuery 1.0 and XPath 2.0		✓	
<b>Large database support</b> Supports very large databases exceeding the capacity of physical memory		✓	
<b>Synchronization</b> Synchronize data bi-directionally with a publish-subscribe model in a sometimes connected network environment			✓
<b>Oracle database caching</b> Provide updateable cache for Oracle 10g backend system	✓		
<b>Small footprint</b>		✓	✓

**Afbeelding 4:** Vergelijking tussen TimesTen, BerkeleyDB en Oracle Lite Edition.

gebruikte technologie. Omnidex maakt onder andere gebruik van quick counts, and/or/not, relationale logica, reeksen, 'sounds like'-methoden, 'partial keys', 'pattern matches', 'table joins', 'composite keys', 'case insensitivity' en 'grouping'. In feite past het dus de aanpak van de bekende zoekmachines op het web toe op de data die vastliggen in databases, HTML, XML, flat files en ASCII-documenten.

## Omnidex wordt gepositioneerd als uiterst schaalbare zoekmachine voor databases

Interessant is ook het fenomeen 'instant counts'. Omnidex wordt ook wel een 'count engine' voor grote databases genoemd. In feite vervangt Omnidex relatief trage scans van duizenden tabellen met geïnverteerde list indexen. Ook andere voorzieningen als keyword access, multi-dimensional access of bijvoorbeeld text searches worden in vogelvlucht op de site behandeld. Belangrijke toepassingsgebieden voor deze technologie zijn onder andere database publishing, het ontsluiten van repository's vol transactionele data, e-business toepassingen, het ontsluiten van content rich data en database marketing.

### Alterian

Alterian is een Britse firma die het al jaren goed doet in een zeer gespecialiseerd segment van de database-markt: het hiervoor reeds genoemde wereldje van de database marketing. De software van het bedrijf luistert naar de naam 'Analytics Led Integrated Marketing', kortweg de Alterian Marketing Suite.

Het bedrijf bestaat sinds 1997 en richt zich uitsluitend op database marketing en aanpalende vakgebieden als 'customer insight' en 'marketing execution'. De technologie die hierbij wordt gebruikt is gebaseerd op de zogeheten Alterian Engine, die volgens het bedrijf in staat is om snel en flexibel zeer grote hoeveelheden data te doorzoeken. Hier bovenop staat een reeks van marketing-applicaties die de engine gebruiken om nuttige informatie uit de almaar uitdijende databases vol klantgegevens te halen. Die applicaties richten zich op het selecteren, visualiseren, plannen en analyseren van marketing-activiteiten.

Dit is bovendien de enige van de vier database-bedrijven die namen van klanten noemt. Bekende gebruikers zijn onder andere Orange, Kodak, Heinz, Merrill Lynch en Vodaphone. De site noemt geen Nederlandse klanten. De lijst met partners daarentegen is indrukwekkend lang en bevat 59 namen die leest als een 'who is who' in de wereld van analytische en marketing software. Over de onderliggende technologie laat het bedrijf weinig los. Wel is duidelijk dat een tool genaamd 'Alterian Data Discovery and Visualisation' (DVV) voor gebruikers een hoofdrol speelt. In feite is dit een query tool waaraan direct al een aantal rapportage-mogelijkheden is gekoppeld. Data access en visualisatie in één, als het ware. Query's kunnen worden gerapporteerd via hulpmiddelen als 'cross tabulation', Venn-diagrammen, mapping, clustering, auto comparison en bijvoorbeeld Pareto. Is een interessante set van data gevonden dan kan deze naar een bestand worden geschreven, zodat export naar een andere applicatie mogelijk wordt. Ook biedt het tool mogelijkheden om data om te zetten in andere formaten, coderingen te veranderen in voor de mens begrijpelijke waarden, terwijl ook attributen kunnen worden toegekend.

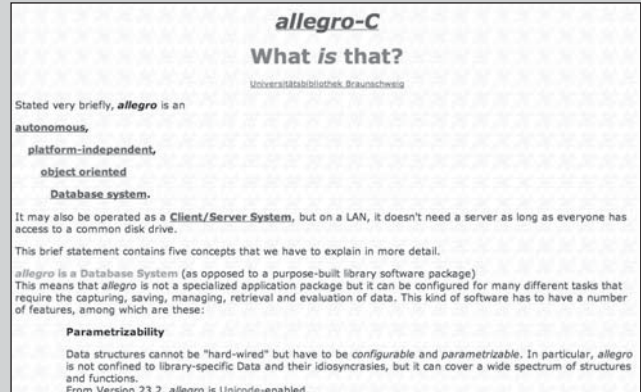


**Afbeelding 5:** Homepage van de DISC, de ontwikkelaar van Omnidex.

# Allegro-C en Netezza

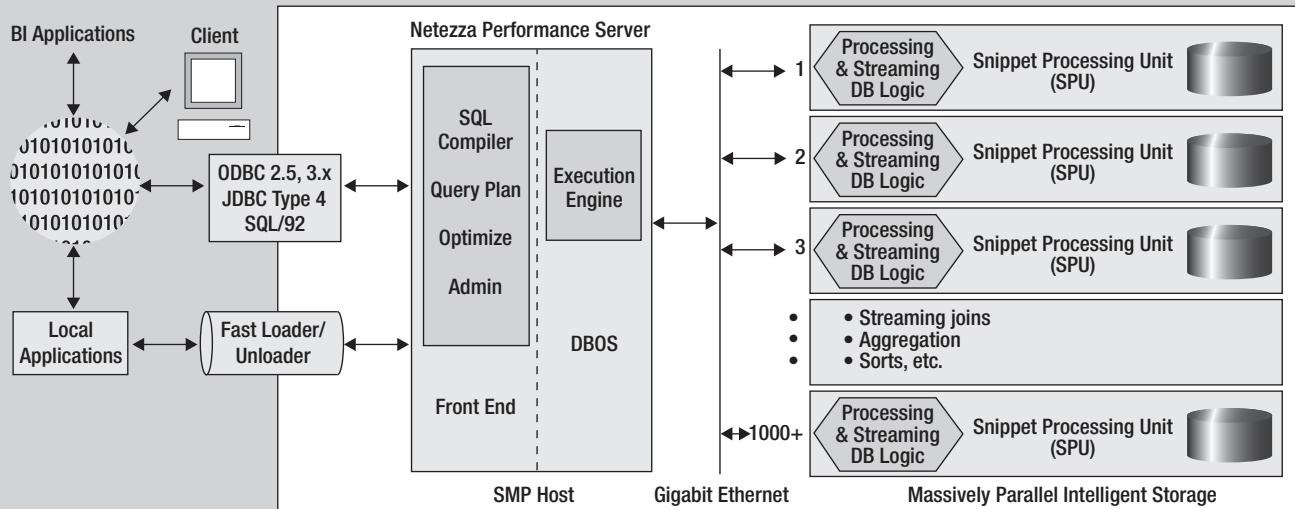
Wel eens van Allegro-C gehoord? Het is een 'autonomous, platform-independent, object-oriented database system' en wordt ontwikkeld aan de Universiteit van Braunschweig. Het kan gebruikt worden in een client/server-opstelling, maar mooier is wellicht de mogelijkheid om deze programmatuur in een netwerkgeving te gebruiken. De enige voorwaarde die dan wordt gesteld, is dat iedere gebruiker toegang heeft tot een gemeenschappelijke diskdrive. Allegro-C is vooral een onderzoeksproject.

Netezza is eveneens interessant. Dit bedrijf bouwt een zogeheten 'data-warehouse appliance'. Neem een database, voeg die samen met een server en de nodige storage en we hebben een Netezza Performance Server. Het voordeel van deze aanpak is volgens het bedrijf een tien tot honderd keer hoger prestatieniveau dan bij een traditioneel datawarehouse. Het verschil zou met name zitten in de optimale afstemming tus-



Afbeelding 6: Verklaring op de website van allegro-C.

sen enerzijds de verwerkingscapaciteit en de toegang tot de opslagruimte en anderzijds de in eigen beheer ontwikkelde manier van bevragen.



Afbeelding 7: De opbouw van de Netezza Performance Server.



Afbeelding 8: Opbouw van de Alterian Marketing Suite.

## Focus op doelgroepen

Alterian is een mooi voorbeeld van een database-product dat geheel op een specifieke doelgroep is gericht. Vandaar ook dat op de website van het bedrijf alle informatie volledig geschreven is met de belevingswereld van die doelgroep in het achterhoofd. Dat is dan ook de kracht van drie van de vier hier genoemde database-producten (Alterian, Omnidex en TimesTen): vergeet de functionaliteit die een algemeen toepasbare database nodig heeft, maar richt alle aandacht en energie op het ontwikkelen van die vaak uiterst gespecialiseerde functionaliteit die alleen voor die ene doelgroep interessant is. Mooie producten zijn het, maar dan alleen voor hun eigen specifieke gebruikersgroep. En RapidBase? Tja, van RapidBase kan eigenlijk alleen maar worden vastgesteld dat er over gesproken wordt. Of – en zo ja waar – het product nog te krijgen is, is tijdens deze zoektocht onduidelijk gebleven.

Robbert Hoeffnagel is freelance journalist.