

Nieuwe databasetechnologie voor datawarehousing (5)

Alterian Engine, de blackbox databaseserver

Rick F. van der Lans

In deze reeks over nieuwe technologie voor datawarehousing is het de beurt aan Alterian Engine. Een SQL-databaseserver met in-memory-achtige kwaliteiten en speciale indextypen, die de gegevens deels kolomgeoriënteerd op schijf vastlegt. Deze wijze van opslag past goed bij het soort vragen dat datawarehouses moeten verwerken.

In onbekendheid doet Alterian Engine niet onder voor de andere in deze serie besproken databaseservers. Maar wat is er zo speciaal aan Alterian Engine? Het combineert enkele in-memory faciliteiten van TimesTen met de speciale indexen van Omnidex. Deze combinatie maakt dat Alterian Engine een databaseserver is die uitstekend ingezet kan worden om de performance van query's op het datawarehouse te versnellen.

In tegenstelling tot bijvoorbeeld Omnidex heeft Alterian Engine geen andere databaseserver nodig. Zij kan volledig op zichzelf functioneren. Hierdoor concurreert het product direct met klassieke relationele databaseservers, zoals DB2, Oracle en SQL Server.

ARCHITECTUUR

In een van de vorige artikelen in deze serie bespraken we TimesTen, een 'echte' in-memory databaseserver. Alterian Engine ondersteunt veel van de in-memory karakteristieken van TimesTen, maar niet alle. Alterian Engine kent bijvoorbeeld geen T-tree indexen. Indexen worden in de database op schijf opgeslagen, net als bij klassieke databaseservers.

De programma-interface is gelijk aan die van elke andere SQL-databaseserver. SQL-instructies, zoals INSERT, UPDATE, DELETE en INSERT, worden via ODBC doorgegeven. De meeste in de ANSI/ISO SQL2-standaard beschreven SQL-instructies worden ondersteund, inclusief CREATE en DROP TABLE, CREATE en DROP INDEX en GRANT. Omdat Alterian Engine een redelijk compleet SQL-dialect biedt, kunnen we relatief eenvoudig BI-tools van onder andere Business Objects en Hyperion via de ODBC-interface met Alterian Engine laten communiceren.

Globaal maken drie aspecten van Alterian Engine het product bijzonder. Alle drie zijn verstopt onder de motorkap. In de eerste

plaats worden gegevens bij het opstarten van de databaseserver zoveel mogelijk in intern geheugen geladen; zie de stappen 1 tot en met 3 in figuur 1. Deze gegevens worden uit de database van Alterian Engine gehaald. De vervolgens op Alterian Engine afgevoerde SQL-instructies worden zoveel mogelijk verwerkt zonder I/O uit te voeren; stappen 4 tot en met 8. De beoogde performancewinst

Alles gebeurt onder de motorkap, en dat heeft geen invloed op de wijze van formulering van SQL-instructies

wordt dus bereikt doordat de schijfeenheden niet benaderd hoeven te worden voor het raadplegen van de gegevens. Ze worden direct uit intern geheugen gehaald. We zullen aan dit aspect niet veel aandacht besteden, omdat op dit gebied de technologie van Alterian Engine veel overeenkomsten vertoont met die van TimesTen.

Ten tweede kent het product een groot aantal verschillende soorten indexen. De meeste indextypen zijn ontworpen om de performance van typische OLAP-query's te versnellen. Hierover later in dit artikel meer. Een groot aantal indextypen is door de oprichters uitgevonden, en sommige zijn zelfs gepatenteerd.

Het product heeft reeds in projecten bewezen met grote hoeveelheden gegevens om te kunnen gaan. De grootste tot nu toe met Alterian Engine gebouwde database is 4 TB groot. Het product

'Query from heaven'

Een van de grootste technologische uitdagingen van het bouwen van een datawarehouse is het realiseren van een acceptabele performance. Anders geformuleerd: hoe transformeren we de 'query from hell' in een 'query from heaven'?

Een mogelijke oplossing is het inzetten van nieuwe databasetechnologie of nieuwe databaseservers. Deze serie behandelt enkele van de alternatieve mogelijkheden. TimesTen, Rapidbase en Omnidex zijn al aan bod gekomen. Nu kijkt Rick van der Lans naar Alterian Engine.

Rij-identificatie	KLANT-NUMMER	NAAM	PROVINCIE	TYPE	...
1	85	Jansen	Zuid-Holland	A	...
2	132	Pietersen	Gelderland	A	...
3	987	Karelsen	Zuid-Holland	B	...
4	56	Borghout	Zeeland	B	...
5	2	Knapen	Gelderland	B	...
6	567	Verbaan	Utrecht	A	...
...	

TABEL 1: EEN KLANTEN-TABEL.

is dus inzetbaar als databaseserver voor een datamart, maar ook voor een centraal datawarehouse.

INDEXEN, INDEXEN EN NOG EENS INDEXEN

De aanwezigheid van de indexen maakt Alterian Engine ook geschikt voor datawarehousing. Helaas valt hierover weinig te melden, want dit houdt Alterian geheim. Hoe het allemaal werkt, is het gezamenlijke geesteskind van de drie oprichters.

Wel zijn een paar algemene beweringen te maken. Nadat tabellen met behulp van CREATE TABLE-instructies gecreëerd zijn, kan de databasebeheerder aangeven op welke kolommen indexen gedefinieerd moeten worden. Dit geschiedt op dezelfde wijze als

Rij-identificatie	KLANTNUMMER	NAAM	...
1	85	Jansen	...
2	132	Pietersen	...
3	987	Karelsen	...
4	56	Borghout	...
5	2	Knapen	...
6	567	Verbaan	...
...	

TABEL 2: AANMAAK VAN INDEXEN OP DE KOLOMMEN PROVINCIE EN TYPE MAAKT DE KLANTENTABEL SMALLER.

PROVINCIE	Rij-identificatie
Zuid-Holland	1, 3
Gelderland	2, 5
Zeeland	4
Utrecht	6
...	...

TABEL 3: INDEX VOOR PROVINCIE.

TYPE	Rij-identificatie
A	1, 2, 6
B	3, 4, 5
...	...

TABEL 4: TYPE-INDEX.

bij elke andere klassieke databaseserver, namelijk met een CREATE INDEX-instructie. Het indextype kan hierbij echter niet ingesteld worden. Alterian Engine zal zelf bepalen welk indextype het meest ideaal is. Dit is afhankelijk van het aantal verschillende waarden in een kolom, de distributie van de waarden, het aantal rijen in de tabel en enkele andere factoren. Alterian Engine kan kiezen uit de bekende B-tree indexen, bitmap-indexen of een van de achtentwintig andere indextypen.

OPSLAAN ALS KOLOMMEN

Elke tabel slaat Alterian Engine op in een apart bestand. Alle waarden waaruit een rij opgebouwd is, worden initieel samen als record opgeslagen. Wat dat betreft dus 'gewoon' recordgeoriënteerd. Maar dit kan veranderen bij het definiëren van indexen. Als een index op een kolom gedefinieerd wordt, kan het systeem de gehele kolom uit het bestand verwijderen en als apart index-bestand opslaan. Of dit gebeurt is afhankelijk van het gekozen indextype. Naarmate men meer indexen definieert, verandert de opslagstructuur langzaam van rijgeoriënteerd in kolomgeoriënteerd.

Veronderstel dat een tabel als tabel 1 reeds gecreëerd is.

Vervolgens creëren we indexen op de kolommen PROVINCIE en TYPE. We weten dat beide kolommen slechts een klein aantal verschillende waarden bevat. Dit zal voor Alterian Engine het signaal zijn de twee kolommen uit de tabel te verwijderen en er speciale indexen voor te creëren. De volgende situatie ontstaat: de KLANTEN-tabel wordt smaller (zie tabel 2), en daarbij worden twee indexen gecreëerd - tabel 3 en 4.

Veronderstel dat er maar twaalf provincies zijn en slechts twee verschillende TYPE-waarden. In dat geval zijn beide indexen klein. De tabel zelf is ook kleiner geworden, want er zijn tenslotte twee kolommen verwijderd. (Ter informatie: klassieke databaseservers verwijderen een kolom niet uit de tabel als er een index op gedefinieerd wordt.)

Beide acties zullen de performance van instructies zoals de onderstaande versnellen:

```
SELECT  KLANTNNUMMER, NAAM
FROM    KLANTEN
WHERE   PROVINCIE = 'Zeeland'
AND     TYPE = 'B'
```

De versnelling ontstaat ten eerste omdat de indexen klein zijn en dus snel in het intern geheugen geladen en daarna doorzocht kunnen worden. Vervolgens zal het zoeken van de relevante records in de KLANTEN-tabel ook sneller verlopen, omdat ook deze tabel fysiek kleiner is geworden. De term die Alterian hanteert voor deze vorm van opslag is *column-based analytical technology* (CBAT).

Het verwijderen van kolommen en daarmee het verkleinen van de tabellen heeft als bijkomend voordeel dat de totale database relatief klein blijft. Zeker als de betreffende kolom lange waarden bevat, is een rijgeoriënteerde opslagstructuur, waarbij een lange waarde veelvuldig wordt opgeslagen, niet efficiënt. Maar het belangrijkste voordeel van kolomgeoriënteerd opslaan is dat veel query's in een BI-omgeving kolomgeoriënteerd zijn. In een datawarehouse zal eerder de vraag gesteld worden: 'hoeveel verschillende provincies zijn er?' dan: 'geef van elke klant alle kolomwaarden'. De eerste vraag is snel te beantwoorden door de PROVINCIE-index te doorlopen. Of veronderstel dat we van elke klant het geboortejaar hebben en we willen weten hoe oud een klant gemiddeld is. Ook hiervoor hebben we alleen de index op de kolom geboortejaar nodig.

Of een kolom inderdaad uit de tabel verwijderd wordt, is afhankelijk van het indextype, wat weer afhankelijk is van een aantal factoren, zoals het aantal verschillende waarden. Bevat een kolom vele verschillende waarden, zoals NAAM, dan blijft de kolom in de tabel behouden. Maar wat Alterian Engine ook beslist, alles gebeurt onder de motorkap en het heeft geen invloed op de wijze waarop de SQL-instructies geformuleerd moeten worden.

WIE ZIJN DE CONCURRENTEN?

Wie zijn nu de concurrenten van Alterian Engine? Uiteraard is elke leverancier van een in een datawarehousesysteem inzetbare SQL-databaseserver een potentiële concurrent. Vanuit technologisch stand-

De historie van Alterian

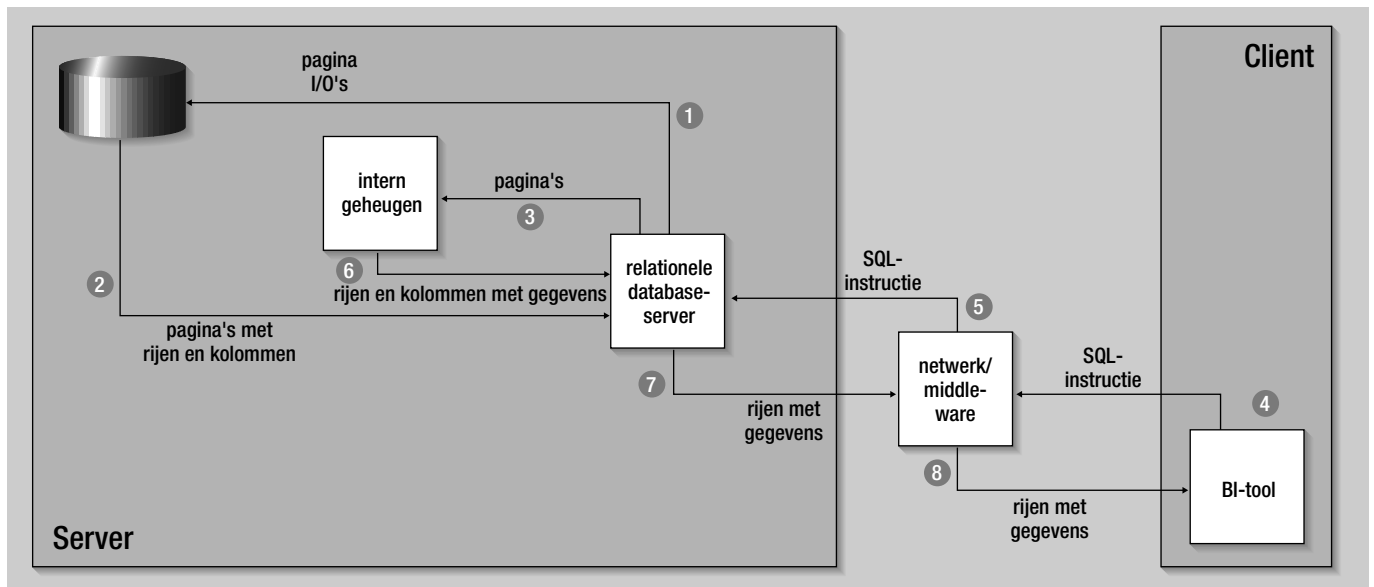
Wat krijg je als drie jonge gasten die, als hun leeftijdgenoten nog proberen de middelbare school af te ronden, al miljoenen hebben verdiend aan het schrijven en verkopen van high-tech software-spelletjes, besluiten om een databaseserver te ontwikkelen? In ieder geval niet een doorsnee oplossing die lijkt op een van de vele, uit Californië komende databaseservers. Je krijgt dan Alterian Engine, een in-memory databaseserver die speciaal is ontwikkeld voor business intelligence-toepassingen.

De oprichters zijn drie ware whizzkids: David Eldridge, Mike Talbot en Tim McCarthy. Eldridge begon zijn eerste softwarebedrijf toen hij veertien was. Mike Talbot lanceerde zijn eerste softwarespel op vijftienjarige leeftijd. Later werd dit spel nog gevolgd door veertig andere spelletjes in een tijdspanne van tien jaar. Ook McCarthy was jong toen hij begon. Op zijn vijftiende studeerde hij af in wiskunde en hij ontwikkelde zich al snel tot programmeur van computerspelletjes.

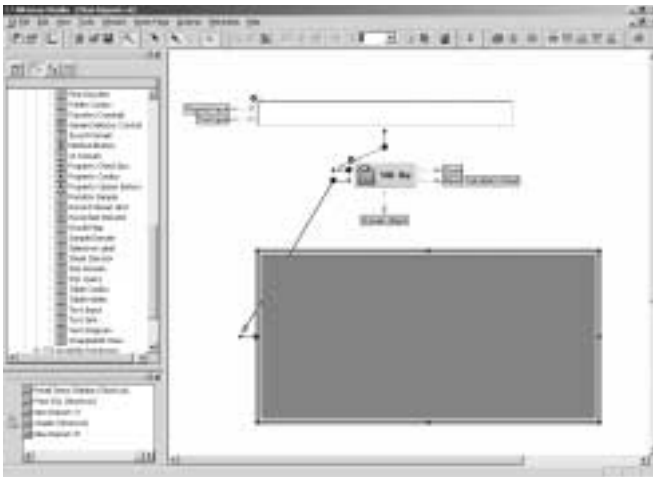
Het bedrijf Alterian is opgericht in 1997. De eerste versie van Alterian Engine kwam in 1998 op de markt. De huidige versie is 2.3 en versie 3.0 zal in de tweede helft van 2002 verschijnen. De producten van Alterian worden niet door de leverancier zelf, maar via partners verkocht. Tevens is Alterian Engine al binnen verscheidene producten geïmplementeerd en wordt het als OEM-product verkocht.

punt echter opereren slechts enkele leveranciers in dezelfde markt.

Alterian Engine is geen multidimensionale databaseserver (Molap-product), zoals Hyperions Essbase en Applix' iTM1. Deze producten hebben geen SQL-achtige interface en gebruiken geen tabelstructuren om gegevens op te slaan. Alterian slaat ook geen geaggregeerde gegevens op, wat normaliter het laadproces vertraagt.



FIGUUR I: ALTERIAN ENGINE IS AAN DE BUITENKANT EEN 'NORMALE' SQL-DATABASESERVER.



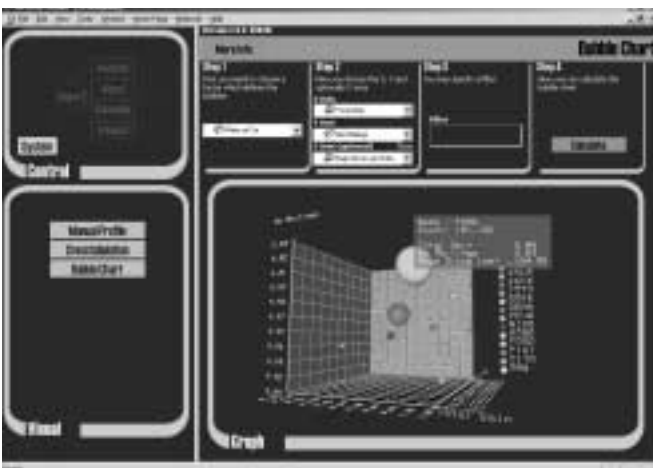
FIGUUR 2: EEN APPLICATIE IN ONTWIKKELING MET ALTERIAN DEVELOPER.

Door de uitgebreide indexmogelijkheden behoort het product ook niet tot de groep van klassieke rdbms'en, zoals DB2, Oracle9i en SQL Server. Deze producten zijn oorspronkelijk ontworpen voor het verwerken van transacties en zijn later uitgebreid om te kunnen functioneren in een datawarehouse-omgeving. Alterian Engine is hiervoor speciaal ontwikkeld. Op het gebied van transacties en beschikbaarheid moet Alterian Engine het onderspit delven tegen deze producten.

De concurrenten van Alterian Engine zijn Omnidex en het in een van de volgende artikelen te bespreken Nucleus (van Sand Technology). Dit zijn alle drie databaseservers die een SQL-interface hebben en speciaal ontwikkeld zijn voor datawarehousing.

ALTERIAN DEVELOPER

Hoewel deze reeks artikelen databaseservers en databasetechnologie behandelt, maken we enkele opmerkingen over de bij Alterian Engine meegeleverde BI-ontwikkelomgeving: Alterian Developer (voorheen Alterian Studio). De meeste BI-tools hebben een door de leverancier bepaalde interface. Uiteraard bepaalt het soort gege-



FIGUUR 3: EEN AFGERONDE APPLICATIE.

vens waarmee gespeeld wordt wat de gebruiker te zien krijgt, maar de globale interface staat vast. Alterian Developer is anders. Het is een soort Visual Basic-achtige ontwikkelomgeving, speciaal bedoeld voor het bouwen van BI-applicaties. Naar eigen inzicht kan een ontwikkelaar allerlei visuele objecten in een leeg window plaatsen en die vervolgens onderling koppelen. Deze objecten kunnen simpele buttons zijn of invoervelden, maar ook driedimensionale grafieken, CHAID-analyses (*chi-squared automatic interaction detection*) voor datamining en zelfs kant en klare objecten voor campaign management.

Ter illustratie tonen we twee schermafdrucken. In figuur 2 is een applicatie in ontwikkeling zichtbaar en figuur 3 toont een reeds afgeronde applicatie. Als alle objecten op de correcte plaats staan, genereert Alterian Developer een Visual Basic-applicatie. Visual Basic-ontwikkelaars zullen geen moeite hebben zich dit product eigen te maken.

SAMENVATTING

De sterke punten van Alterian Engine zijn de volgende.

- We hebben te maken met een volwassen product, dat samen met Alterian Developer een sterke combinatie biedt met unieke mogelijkheden voor het ontwikkelen van BI-applicaties.
- De geboden indextechnologie is van hoge kwaliteit en zal vele BI-omgevingen een performanceverbetering geven.
- Dankzij de SQL-interface zullen ontwikkelaars met kennis van andere SQL-producten niet veel tijd nodig hebben om het product te bestuderen.
- Veel van de technologie is verstopt onder de motorkap, waardoor databasebeheerders weinig tijd hoeven te steken in indexbeheer.

Alterian Engine kent ook enkele minpunten.

- De leverancier geeft zelf aan dat de beperkingen van het product niet zozeer liggen in de databasegrootte, maar in de hoeveelheid gebruikers. Eén installatie kan maximaal dertig tot veertig gebruikers aan.
- De software is alleen beschikbaar voor Windows-platforms.
- Helaas verstrekt de leverancier niet veel informatie over die technologisch interessante binnenkant. Deze geheimhouding zal toch enkele klanten weerhouden om tot aanschaf over te gaan. Alterian zou er goed aan doen meer openheid te verschaffen, waardoor het voor klanten duidelijker is wat de potenties van dit product zijn.
- Omdat de leverancier de distributie niet zelf regelt, maar dat aan derden overlaat, moet de koper de kwaliteit van de lokale ondersteuning bestuderen. ●

Rick F. van der Lans (rick@r20.nl) is onafhankelijk consultant en redactieadviseur van Database Magazine.