



Value-based database biedt unieke manier om data op te slaan

illuminate correlatie-database

Bram Dons

Wat ondernemingen nodig hebben is niet een eenvoudiger BI, maar een betere BI. Er is recentelijk een nieuw product op de markt verschenen dat een bruikbaar alternatief kan bieden voor de SQL-gebaseerde BI-systemen.

Fysieke datastructuren in relationele databases zijn doorgaans gebaseerd op een compromis tussen de optimalisatie van prestaties en een perfect, in een zo hoog mogelijk norm genormaliseerd, entity/relatiemodel. Zoals bekend worden in een relationele database data in tabellen opgeslagen die records bevatten, waarbij elk record een aantal velden bevat. Deze zijn gekoppeld via foreign key relaties en geïndexeerd om de toegangssnelheid te optimaliseren.

Echter, alles rondom deze structuur is gebaseerd op compromissen tussen computerprestaties en bruikbaarheid. Elke ontwerper van relationele databases weet dan ook dat een goed database-schema uiterst belangrijk is, omdat het schema-ontwerp het bereik en de limieten definieert van de informatie die straks uit de database kan worden opgevraagd. Dit geldt vooral bij datawarehouses, analytische databases en andere informatie-systemen. Het is dan ook geen overdrijving om te stellen dat het ontwerp van het informatieschema de toekomstige mogelijkheden van de onderneming definieert en tegelijkertijd beperkt. Veel zakelijke managers zijn zich waarschijnlijk niet bewust van de toekomstige invloed die een nieuw database-ontwerp heeft op de informatievoorziening van de onderneming en laten dit volledig over aan de competentie van de IT-afdeling.

Op *executive* niveau zijn ook maar weinigen geïnteresseerd of getraind in het ontwerp van een relationeel database-schema (waarom zouden ze ook), en vaak is men er niet van op de hoogte dat alle bruikbare relationele databases een goed ontworpen schema vereisen, zodat die later kan dienen als waardevol analytisch instrument voor het analyseren van bedrijfsinformatie. De oplossing voor dit probleem zou kunnen zijn om de zakelijke managers te trainen in het ontwerp van databases of om, meer realistisch, de mystiek en magie van het relationele database-ontwerp weg te nemen. De structuur van de analytische data-bases zou het complete probleem van schema-ontwerp

moeten wegnemen. Maar waarom wordt de in de relationele databases opgeslagen informatie niet geabstraheerd en gebruikt in een ander soort, niet op SQL gebaseerd, query tool? Op dat idee kwamen de ontwerpers van de 'value-based' database en kan het een mogelijk antwoord zijn op het hiervoor beschreven probleem. Of dat ook in de praktijk kan worden waargemaakt zien we in een evaluatie van dit nieuwe product van de firma illuminate.

Value-based storage systeem

Fysieke data-structuren in relationele databases zijn dus een compromis tussen prestaties en een perfect geoptimaliseerd entity/relatiemodel. De kolomgeoriënteerde databases bieden een alternatief voor deze structuur, maar hebben weer te maken met zaken als de laadsnelheid bij hoge volumes en met de uitvoering van recordgeoriënteerde processen als export en rapportage van volledige records.

In de radicale nieuwe benadering van de firma illuminate wordt de fysieke datastructuur geheel vanuit een externe informatie-structuur geabstraheerd. Data values worden zodanig opgeslagen dat er een optimaal werkend opslagsysteem ontstaat. Er bestaat geen fysiek schema in dit systeem, alleen een logisch schema dat de verschijningsvorm definieert voor gebruikers en programmatische toegang. Alle unieke data worden eenmalig opgeslagen, ongeacht locatie en frequentie. Data values worden per datatype opgeslagen op een zodanige manier dat deze zijn geoptimaliseerd voor het beheer van raw data values, in plaats van de optimalisering voor gebruikers-query's in tabelstructuren. De consequentie daarvan is dat elke toegang tot de opgeslagen unieke data altijd met behulp van een index plaatsvindt. De dictionary bevat geavanceerde metadata die betekenis geven aan de opgeslagen raw data en ondersteunt autonavigatie tussen de logische tabellen in de database. De value-based database

heeft extern de verschijningsvorm van een relationele database maar met additionele features en ongewone prestatiekenmerken. De fysieke database is aanzienlijk kleiner dan een relationele of kolomgebaseerde database. Door de ongebruikelijk toegepaste compressietechniek verbetert de compressiefactor naarmate de database groeit. Zo kan een kleine database tweemaal zoveel ruimte innemen als de in de database geïmporteerde raw data in beslag nemen. Bij twee miljoen opgeslagen records is de opslagruimte ongeveer even groot als de geïmporteerde raw data, maar bij een grotere database wordt de ruimte minder dan de raw data innemen.

Het opvragen voor analytische doeleinden verloopt een stuk sneller dan elke andere relationele database en de recordgeoriënteerde operaties zijn ook sneller dan de kolomgeoriënteerde databases. Bovendien worden nieuwe type query's ondersteund die niet mogelijk zijn met de conventionele database-systemen. De value-based database biedt een unieke manier om data op te slaan die een aantal belangrijke voordelen voor BI processing leveren.

illuminate database-structuur

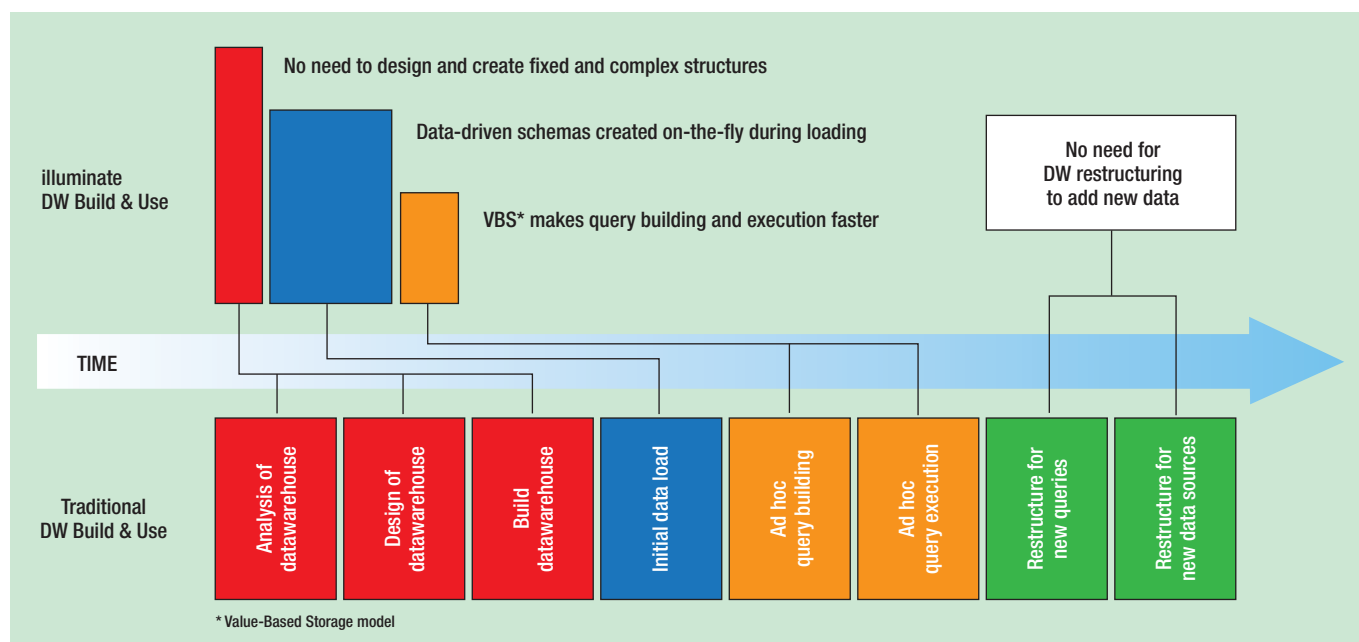
Het illuminate geabstraheerde database-model bestaat uit drie primaire fysieke sets met opgeslagen en beheerde objecten: een data dictionary (metadata); een index en linking data set (additionele metadata) en de eigenlijke data values met de opgeslagen informatie. De data value storage bevat een instance met unieke data values die werden gedetecteerd tijdens alle geïmporteerde data streams. Dat houdt in dat het systeem altijd het hoogste normalisatieniveau kent: het niveau van unieke waarden. Dit elimineert de noodzaak om data sets in een logisch schema te normaliseren, met als resultaat een structuur die eenvoudig door iedereen kan worden begrepen. De data values

worden in een geordende set datatypes opgeslagen, alle integerwaarden in een set, karakters in een andere, enzovoort. Deze opslagarchitectuur optimaliseert in hoge mate de verwerking van value-based data. Door de data value-abstractie, met de betekenis in logische records, vervalt de noodzaak om complexe ordening en linking sets te bouwen, zoals de meeste kolomgeoriënteerde databases doen. De bij de value-based database toegepaste relaties bieden een soortgelijke functie als het foreign key concept bij de relationele database, maar in dit geval is de relatie in de dictionary bekend en opgeslagen als een data value. Deze database-structuur maakt dat de navigatie tussen logische tabellen volledig automatisch verloopt. Bijvoorbeeld, om alle klanten te vinden die een bepaald product kopen ziet een query in de 'gesimuleerde' SQL er in een value-based database als volgt uit:

```
select klanten.naam where bestelling.code = 123
```

Het geabstraheerde database-model bestaat uit drie primaire fysieke sets

Op basis van een standaard foreign key zouden voor hetzelfde SQL statement in een relationele database meerdere joins nodig zijn voor de complete selectie. Natuurlijk kan het foreign key proces worden toegepast in de illuminate database als een link, anders dan de gedefinieerde inter tabel links die zijn vereist voor een speciale query. Omdat ook alle data zijn geïndexeerd kan



Afbeelding 1: Value-based Storage Model.

elk veld een foreign key en relaties bevatten die niet vooraf door de DBA zijn voorzien en kunnen dan ook snel door de gebruiker worden ingevoerd. De data dictionary bevat de doorsnee metadata, plus additionele statistische data over tabellen, kolommen en occurrence values in het logische schema. Het bevat ook informatie over de relaties tussen de logische tabellen.

De index en linking storage omvat alle data die wordt gebruikt om de inhoud van een record vanuit de ordered set values te lokaliseren.

De illuminate database is geen RAM-gebaseerd opslagsysteem, maar maakt wel van zoveel mogelijk RAM gebruik als het operating systeem te bieden heeft. Net zoals bij grote relationele databases zal ook een value based database met meer RAM beter presteren. In het algemeen is 2 GB RAM voldoende voor een optimale toegang tot 100 miljoen records, 4 GB is genoeg voor een database die tien keer groter is. Huidige door illuminate uitgevoerde testen geven aan dat 6 GB RAM voldoende is om een database te ondersteunen met 20 miljard records.

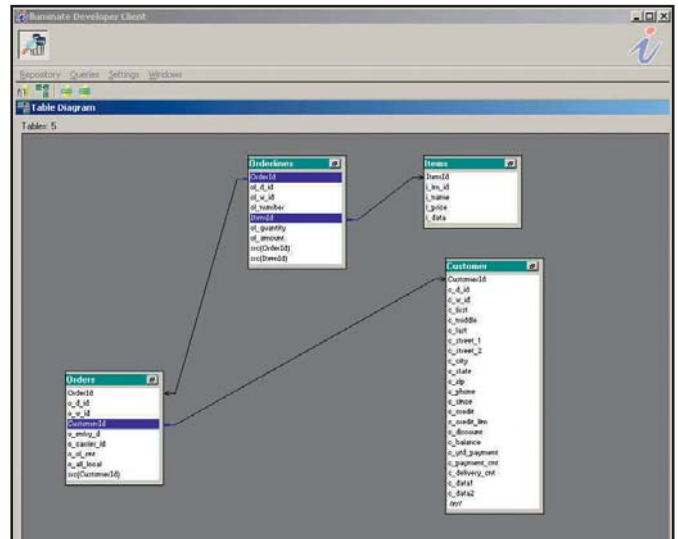
illuminate installatie

De illuminate software bestaat uit een client- en server-gedeelte. De gebruiker communiceert met de illuminate Engine Service via de illuminate Explorer, een standaard user interface genaamd MDI. Na het aanmaken van de illuminate database en repository kan worden begonnen met de import van data uit de externe databases.

Voor het opzetten van de database is het niet nodig om na te denken over de indexering bij de opbouw van de database, de tabel en kolomnamen zijn eenvoudig aan te passen, segmentaties en berekeningen zijn snel en eenvoudig toe te voegen. De datavoorbereiding wordt hierdoor verkort en door de eenvoud is het niet altijd nodig om segmentaties door de IT-afdeling te laten voorbereiden.

6 GB RAM is voldoende om een database te ondersteunen met 20 miljard records

Voor onze test importeren we een paar tabellen uit een Oracle 11g en Microsoft SQL 2005 database. Voor de verbinding met beide databases wordt allereerst via Microsoft's ODBC Data Source Administrator een ODBC connection geconfigureerd met beide databases. In het illuminate 'Select an ODBC font to connect' menu zien we de ODB-verbindingen terug. In het daaropvolgende menu 'Data Selection' kan een SQL tabel worden geselecteerd. Tevens zien we de toegepaste SQL query voor het opvragen van de tabel. Eventueel kan men de SQL query aanpassen als bepaalde kolommen niet nodig zijn, kolommen hernoemd moeten worden, extra selectiecriteria toegevoegd moeten worden of voor het toevoegen van extra berekeningen.



Abbeelding 2: Table Diagram menu.

In het uit drie hoofdonderdelen bestaande Metadata Explorer menu zien we de vijf geïmporteerde tabellen in boomstructuur weergegeven vorm terug. Klikkt men op 'Table' dan zien we een menu met een aantal opties. In 'Statistics' zien we statistische gegevens over illuminate en de Repository.

Het Table Diagram toont een grafische presentatie van de tabellen en de relaties daartussen. Het Table Diagram bestaat uit drie objecten: tables, columns en relationships. Elke in de Repository gedefinieerde tabel wordt als een object in het Table Diagram afgebeeld en kan op verschillende manieren worden weergegeven, in elkaar geklapt met alleen de tabelnaam, of uitgeklast waarbij alle kolommen zichtbaar zijn. Kolommen worden op tabelnaam afgebeeld en geformatteerd per type. Relaties worden gedefinieerd tussen instances op basis van een relatietype kolom. Het diagram laat het onderscheid zien tussen de 'transmuted' relaties en degene die dat niet zijn. Wanneer een relatie is 'transmuted' dan verschijnt de relatiepijl in het zwart en verbindt deze de source column met de target column in de relatie. Voor niet-transmuted relaties wordt de pijl in geel afgebeeld. Het merendeel van de acties die Repository biedt is via dit menu uit te voeren, zie afbeelding 2.

Query Window

Met behulp van het Query Window menu zijn interactief query's te creëren en te modificeren. Het ontwerpproces bestaat uit de toevoeging van nieuwe condities aan de query en het resultaat van de query zichtbaar maken. Het Design tabblad laat de objecten in de Repository zien die voor een query kunnen worden gebruikt, waarbij men een keuze kan maken uit de beschikbare tabellen, kolommen en attributen. De query's kunnen worden voorzien van de gebruikelijke operators als Equal to, Less than, enzovoort. Het resultaat van een query in illuminate wordt samengesteld uit de geselecteerde instances, onafhankelijk van de tabel waartoe ze behoren.

In het Query's Diagram zien we een grafische weergave van alle

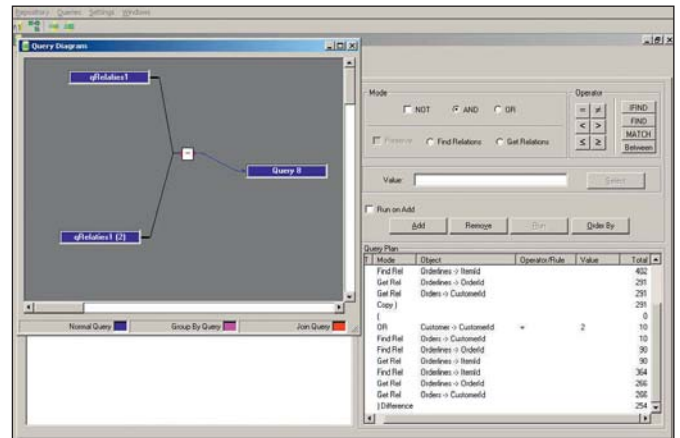
actieve query's en het biedt de mogelijkheid om nieuwe query's te creëren op basis van al bestaande query's. In het diagram zien we alle actieve query's afgebeeld en de relaties daartussen waarbij een kleur het type query aangeeft. Tussen query's onderling kan een Union-, Intersection-, Difference-, Correlation- of Join-operatie worden uitgevoerd. Het resultaat van de gecombineerde acties wordt in het menu afgebeeld met verschillende symbolen. Tenslotte, query's kunnen ook vanuit een extern ODBC-gekoppeld database-tool worden aangeroepen, zie afbeelding 3.

Analyse-optie

De data analyse-optie maakt het mogelijk om automatisch alle samengestelde data uit de verschillende tabellen of expressies op te vragen. De data worden in een of meer verschillende kolommen of expressies weergegeven, die zijn berekend op basis van een bepaalde operatie (counter, add, average, maximum of minimum). De gebruiker hoeft slechts deze kolommen en expressies die men wil analyseren te verslepen, zonder te weten waar ze vandaan komen of hoe ze onderling zijn gekoppeld. De analysis tool kan het resultaat in twee formats laten zien, tekstueel en grafisch.

Conclusie

illuminate is een tool waarmee data uit verschillende typen databases zijn te extraheren. De gebruiker kan daarna zelf de tabellen koppelen, waarmee een database ontstaat die tegemoet komt aan de zakelijk gestelde eisen voor data analyse. Door de gebruikersvriendelijke interface is de inleertijd voor het samenstellen van de database, query's samenstellen en de uitvoering van een data analyse zeer kort. De zakelijke gebruiker of marketeer hoeft daarbij geen kennis te hebben van de SQL-taal. Het grote voordeel van deze query tool is dat het de gebruiker



Abbeelding 3: Query Diagram menu.

onafhankelijk maakt van de IT-afdeling. De gebruiker hoeft niet langer te wachten totdat de database-expert de gevraagde query heeft ontwikkeld, maar kan zelf aan de slag gaan bij het opvragen en analyseren van database data. De toepassing van de speciale door illuminate ontwikkelde value-based architectuur ontkoppelt de strikte relatiestructuur van de relationele database en biedt de gebruiker de mogelijkheid om zelf snel een database-architectuur te ontwikkelen. De toegepaste index- en pointer-technologie van de value-based database maakt een snelle query mogelijk van de database, een snelheid die met een relationele database nooit gehaald kan worden. Door bovendien de data te abstraheren uit de originele relationele database en deze offline te gebruiken, worden de server en de daarop draaiende databases voor uitvoering van data analyse-query's niet langer belast.

Meer informatie is op www.i-illuminate.com te vinden.

Bram Dons is onafhankelijk IT consultant.

Update

Doe mee aan het Nationaal Datawarehouse Onderzoek 2008

Database Magazine organiseert samen met Atos Origin het Nationaal Datawarehouse Onderzoek 2008 (DWO 2008). Het onderzoek richt zich niet alleen op de IT-afdeling, maar ook op de business als eindgebruiker van het datawarehouse. Deze eindgebruikers kunnen juist inzicht verschaffen in de acceptatie en de ondersteuning die de datawarehouse-omgeving biedt bij het nemen van beslissingen. Voor IT-afdelingen en business-afdelingen zijn aparte vragenlijsten opgesteld. Deelnemers aan het DWO 2008 krijgen:

- een speciale PDF-editie van de gedetailleerde uitkomsten van het onderzoek, de conclusies en aanbevelingen;
- korting op de toegang tot het Datawarehouseing & Business Intelligence congres in het najaar van 2008;
- indien de deelnemer dat wenst, een kostenloze terugkoppeling van zijn/haar specifieke datawarehouse-omgeving. Hierbij worden in een advies verbeterpunten aangegeven.

In Database Magazine worden de resultaten van het onderzoek toegelicht. Daarnaast verschijnt er een uitgebreide rapportage met daarin de uitgangspunten, resultaten, conclusies en aanbevelin-

gen van het onderzoek, dat tijdens het jaarlijkse congres Datawarehousing & Business Intelligence zal worden gepresenteerd.

Lezers van Database Magazine en relevante zusterbladen worden tijdig op de hoogte gebracht wanneer de online enquête ingevuld kan worden. Doe mee en toets uw eigen datawarehouse-omgeving aan het maturity-model! Het DWO 2008 is een project van Bernard Beefink en Sjoerd Janssen, respectievelijk project manager BI/DWH en consultant BI/DWH bij Atos Origin, en Hans Lamboo, hoofdredacteur Database Magazine.