

CloudView flexibel in te zetten naast de database

Invloed van Search-Based Applications op BI

Ronald Baan

BI gaat steeds meer technieken gebruiken die in de wereld van Enterprise Search al jaren gemeengoed zijn. Search-Based Applications (SBA) op basis van de zoektechnologie van Exalead bieden nu al enorme voordelen als snelheid, bredere inzetbaarheid en kostenbesparingen.

Vorig jaar was ik op het BI-event in Hoevelaken op uitnodiging van Exalead, net als dit jaar ook sponsor. Ik werk al ruim 20 jaar met zoektechnologie en ken de oplossingen van onder andere Exalead, Autonomy en Google van binnen en van buiten. Ik verwonderde me over de activiteiten van de traditionele BI-aanbieders die met alle macht proberen om niet-gestructureerde informatie (zoals we tweets, blogs, webpagina's en documenten mogen noemen) in hun oplossing te zuigen en dan ook nog even hun bestaande of aangepaste ETL-tools erop los te laten om allerlei verrijkingen te bewerkstelligen. Ik hoop dat ze dat gaat lukken, uit ervaring weet ik dat dat niet eenvoudig is.

Per definitie is een zoekmachine ook een content integratiemachine

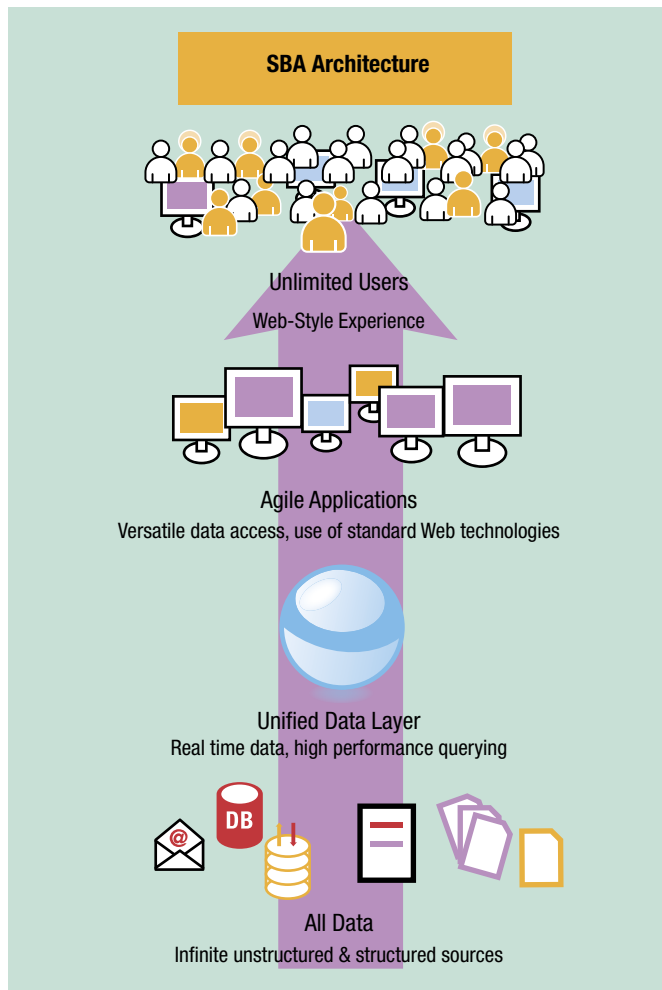
Zoektechnologie leveranciers, zoals Exalead CloudView, zijn hier al jaren mee bezig en hebben daar goede, eenvoudig te beheren ETL-oplossingen voor. Met zoektechnologie is een informatie integratielaag te maken met data en informatie uit allerlei bron-systemen, zoals databases, document managementsystemen, websites, en in allerlei formaten, waaronder database records, XML, PDF, Word, HTML enzovoort. Via de geavanceerde ETL-processen wordt hier een uniforme en universele laag van gemaakt. Deze laag bevat dus genormaliseerde en verrijkte data en documenten.

Op die laag komen allerlei Search-Based Applications. Dit zijn specifieke zoektoepassingen die voor specifieke werkprocessen informatie toegankelijk maken voor bevraging of rapportage. In dit artikel wil ik de mogelijkheden van SBA's ingaan.

Verschillen SBA's met database-oplossingen

SBA zijn anders dan toepassingen gemaakt met database-technieken.

1. SBA's spreken de taal van de gebruiker. De gebruiker hoeft vaak niet eens te weten wat de dataset is of hoe erin te zoeken. Alles is doorzoekbaar en er kan vaak gezocht worden in natuurlijke taal. Hierbij kunnen synoniemen, betere termen, meervouden en werkwoordvervoegingen meegenomen worden. Door de inbreng van experts kunnen alternatieven of verfijningen worden voorgesteld (thesaurus of taxonomie) en fuzzy logic en semantiek door de zoekmachine kunnen de gebruiker verder helpen;
2. SBA's plaatsen informatie in context. SBA's halen hun informatie uit allerlei bronnen en functioneel beheerders bepalen hoe deze informatie het meest betekenisvol kan worden aangeboden. Door deze context krijgt de gebruiker sneller inzicht en overzicht om vanuit het specifieke werkproces actie te ondernemen;
3. SBA's tonen relevante informatie. Op basis van relevantieregels bieden de SBA's een resultaatset aan. Anders dan een set die aan de criteria voldoet, kunnen SBA's een kleinere set geven met de meest relevante data om analyses op te doen of juist een set die niet helemaal aan de criteria voldoet, maar wel relevant is. Gebruikers kunnen hierdoor de *best mogelijke* set gebruiken;
4. SBA's zijn eenvoudig te maken. SBA's zijn gebaseerd op SOA en standaarden als REST, SOAP. Daarbij kunnen de SBA's en de zoeklaag gevoed worden met datasets en ontologie in RDF en OWL. Veel organisaties gebruiken deze manier van applicatieontwikkeling al. SBA's zijn tevens applicaties die kenmerkend mash-ups zijn. Zij kunnen hun informatie overal vandaan halen via webservices en zijn dus gemakkelijk en snel te implementeren.



Afbeelding 1: SBA architectuur.

Zoekmachine om datasets te doorzoeken?

Een zoekmachine is vanaf de grond af ontworpen en ingericht om snel door grote hoeveelheden data en documenten uit allerlei bronnen te zoeken. Efficiënt opslaan van data of beheer hiervan zijn veel minder interessant. Een zoekmachine doorzoekt data en documenten die in andere systemen beheerd worden.

De architectuur van een zoekmachine is erop ingericht om heel efficiënt query's af te handelen. De architectuur van de zoektechnologie is zodanig dat distributie van data en query-afhandeling over meerdere servers om zo datasets tot in de Petabytes en query loads van tientallen per seconde snel te kunnen afhandelen met gemak geconfigureerd kan worden.

Op het internet zien we de giganten als Google, maar ook Exalead (www.exalead.com), die ruim binnen 1 seconde miljarden webpagina's doorzoeken. Door het parallel en geclusterd inrichten van servers kunnen fenomenale responstijden worden gehaald, als het al niet op 1 server kan.

Per definitie is een zoekmachine ook een content integratiemachine. Allerlei niet-gelijksoortige interne en externe bronsystemen worden verwerkt, verrijkt, gekoppeld. Op deze manier kan een klant uit het CRM worden gekoppeld aan een bezoekverslag uit het DMS en samen met de laatste nieuwsberichten over die

klant en het sentiment uit de sociale media worden gepresenteerd. Overigens zijn zoekmachines niet een vervanging voor databases, ze zijn vooral complementair. Voor snelle toegang, kwaliteit van resultaten en relevantie zijn ze ideaal. Voor opslag, beheer en zeer complexe operaties zijn databases meer geschikt.

Meer gebruik, meer waarde

BI breder in de organisatie inzetten zal zeker meer waarde opleveren. BI kan als toepassing worden neergezet voor de BI-experts, echter BI-functies kunnen ook binnen bestaande werkprocessen worden gebruikt door veel meer medewerkers. Een groot deel van de kosten van BI zit in het verzamelen en organiseren van de data. Het toegankelijk maken en analyses mogelijk maken zijn initieel ook behoorlijke posten. Het verder uitbreiden van het gebruik en de hoeveelheid data, zeker met zoekmachines die veel query's per seconden aan kunnen, is relatief goedkoop. Dat maakt de ROI van bredere inzetbaarheid zeer interessant.

Om het gebruik van BI binnen de organisatie breder in te zetten, is het echter wel noodzakelijk dat het gemakkelijker in gebruik wordt. Hieronder wordt een aantal punten genoemd die hieraan bijdragen.

Kennis van en over de organisatie

Zoekmachines kunnen hele relevante terugkoppeling aan de gebruiker geven over de resultaatset. Als die wordt gecombineerd met kennis van en over de organisatie, dan ontstaat er de best mogelijke oplossing. Experts kunnen de zoekmachine voeden met structuren als een ontologie en taxonomie, maar ook met regels voor de verrijking van de informatie. In de SBA's worden daar bovenop de werkprocessen expliciet ondersteund. Door de bundeling van techniek en inhoudelijke kennis wordt het mogelijk om met SBA's de werkprocessen effectiever te maken.

Document- en databeveiliging

Een bredere inzet van data en documenten voor BI houdt ook in dat er meer gebruikers zullen komen die niet overal bij mogen. Aangezien BI data en documenten uit allerlei bronnen haalt, is ook de kans groot dat record- en documentbeveiliging een hele uitdaging wordt. Denk maar eens aan de verschillen in beveiligingsmodellen tussen een netwerkschijf, Oracle database en Documentum DMS.

Naast de complexiteit, mag het toepassen van de beveiliging ook niet ten koste gaan van de snelheid. Dit is vrijwel niet op te lossen door zelfbouw, dit moet echt in de kern van het product zitten. Zoekmachines als Exalead CloudView, die al jaren Enterprise Search oplossingen bieden, hebben hier goede en schaalbare oplossingen voor.

Eenvoud in gebruik

Het gebruik van BI-toepassingen kan veel eenvoudiger voor een groot aantal specifieke toepassingen. Met SBA's is het eenvoudig om veel verschillende toepassingen te maken, welke allemaal op

dezelfde dataset werken. Een aantal zal de hele set gebruiken, vele zullen een specifieke subset gebruiken. Hierdoor kunnen SBA's specifiek voor werkprocessen gemaakt worden. Ze doen en kunnen dan enkel dat waarvoor ze bedoeld zijn. De aandacht is hier minder om allerlei analyses mogelijk te maken. Het gaat veel meer om het mogelijk maken van specifieke analyses, welke door medewerkers zonder training uit te voeren zijn. Visualisaties zijn hier een belangrijk onderdeel, maar ook het doorklikken naar de onderliggende data om verdere analyses uit te voeren en doorzetten van resultaten naar een volgende stap in het proces.

Een groot deel van de kosten van BI zit in het verzamelen en organiseren van de data

Naast het gebruik binnen het werkproces, is zoektechnologie ook in staat om relevante data en informatie via e-mail of andere kanalen onder de aandacht van de gebruiker te brengen. Dit kan gaan om een relevante update van bepaalde data, maar het zou ook een nieuwe versie van een compleet rapport kunnen zijn. Op deze manier wordt het werken met data en informatie verder geautomatiseerd en wordt op een eenvoudige manier meer efficiëntie bereikt.

Ondersteuning van de gebruiker

De 'democratisering' van het gebruik van data begint een hot topic te worden. Steeds meer data komen beschikbaar en de tools voor 'slicing-and-dicing' en visualisaties worden steeds beter en gemakkelijker te gebruiken. Straks kan iedereen het, is het idee.

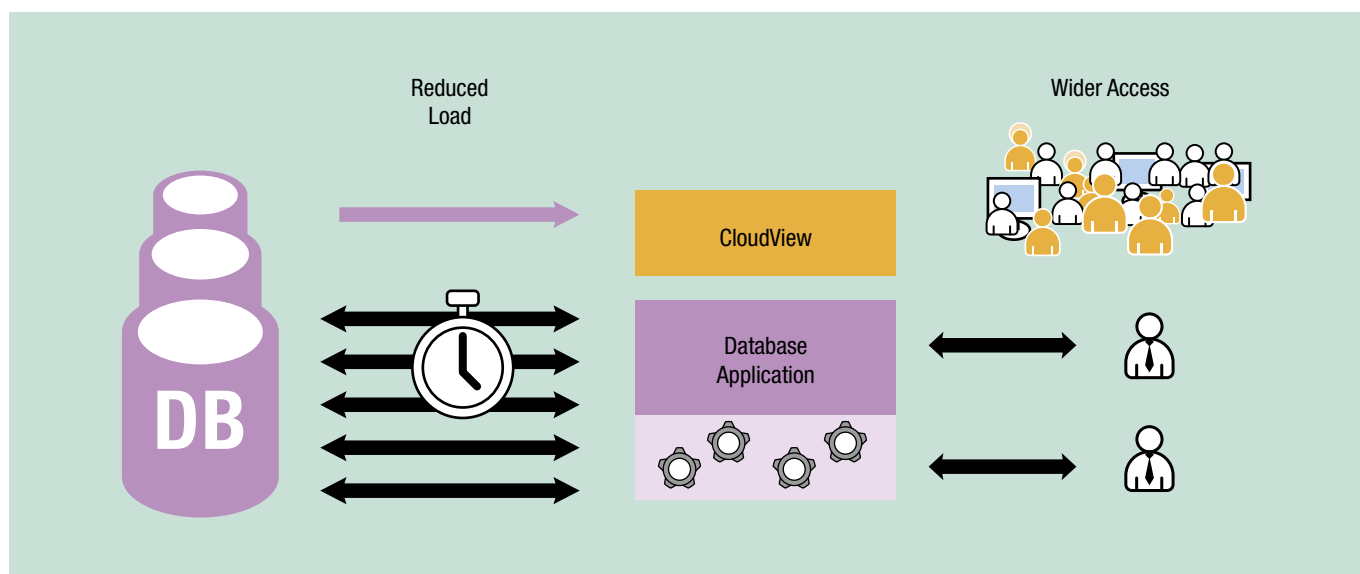
Hier is nog wel goede ondersteuning van de gebruiker voor nodig. Een gebruiker kan bijvoorbeeld met een 'onhandige' eerste query beginnen, waardoor het lastig wordt om nog zinvol verder te selecteren en analyseren. Die onhandigheid kan zitten in een te specifieke query, waardoor er te weinig bruikbaar terugkomt. Een onhandige query kan ook zijn dat het niet goed het onderwerp beschrijft. De gebruiker denkt vaak heel veel, maar geeft slechts een paar criteria in. De minst slechte onhandige query is de te brede query, die teveel teruggeeft, maar waar via terugkoppeling met facetten, gerelateerde termen, onderwerpen en overzichten wel de goede keuzes aan de gebruiker terug te koppelen zijn.

Functioneel beheerders moeten, zoals eigenlijk bij ieder systeem, het gebruik in de gaten houden en gebruikers als groep of individueel bijsturen via de SBA's. SBA's kunnen voor specifieke werkprocessen andere resultaten tonen, suggesties geven van betere selecties enzovoort.

Database offloading

Zoals al eerder vermeld zijn databases voor bevraging niet optimaal. Bij het inrichten van een database moeten er keuzes gemaakt worden om de nadruk op beheer of juist op bevraging te leggen. Bij een zoekmachine is die keuze er niet. Een zoekmachine is er enkel voor de bevraging, maar dat gaat dan ook supersnel. Door het inzetten van zoektechnologie is er de ideale situatie: een database voor beheer van de data en informatie, een zoekmachine voor de bevraging.

Zoals eerder beschreven bestaan de database en zoekmachine naast elkaar. De hoeveelheid hardware die nodig is voor een sub-second performance is voor een database vele malen hoger dan voor een zoekmachine. Extra licenties en beheer tikken behoorlijk door in de totale kostprijs. De TCO van een gecombineerde database-zoekmachine oplossing zal in de meeste gevallen veel gunstiger uitvallen.



Afbeelding 2: Cloudview naast de database.

Gefco – praktijkvoorbeeld uit de logistiek

Gefco is een marktleider op het gebied van geïntegreerde logistiek voor fabrikanten. Gefco vervoert en levert end-to-end supply chain diensten voor industriële klanten in onder andere de automotive, elektronica en detailhandel. Gefco staat in de top 10 van Europa's logistieke groepen met een omzet van € 2,8 miljard in 2008. De 10.000 medewerkers bedienen klanten in meer dan 100 landen op 6 continenten.

De uitdagingen voor Gefco op het gebied van het volgen van voertuigen waren als volgt:

- verhoog de updatecyclus van de data en de query performance;
- lever hoge prestaties op hoge datavolumes: 80 landen, 600.000 voertuigen en 100.000 acties per dag;
- minimaliseer eventuele upgradekosten en zorg dat reeds gedane investeringen in IT beschermd blijven;
- biedt klanten een innovatieve dienstverlening.

De reden hiervoor was dat er met de bestaande infrastructuur en diensten veel succes was geboekt. Daardoor liepen echter het datavolume en het aantal query's zodanig op, dat er performance issues waren en dat het moeilijker werd om nieuwe dienstverlening uit te bouwen. Na de selectie van Exalead CloudView om BI-functies voor het Gefco portaal te leveren, is hun interne IT-dienstverlener een aantal weken aan de slag gegaan.

De 3 Terabyte data konden moeiteloos worden ontsloten met Exalead CloudView en het resultaat was ernaar:

- de updatecyclus is terug van 24 uur naar 15 minuten;
 - query's worden nu afgehandeld in maximaal 2 seconden, tegenover tot wel 30 seconden eerst;
 - de kosten voor gebruik zijn 50 procent van de kosten voor de Oracle database;
 - door het gebruik van SBA kunnen nieuwe rapportages en toepassingen snel worden geïmplementeerd met een goede ROI.
- Gefco is klaar voor de toekomst. Nieuwe leveranciers kunnen door de zoekarchitectuur moeiteloos worden toegevoegd, ook al nemen de data behoorlijk toe en innovaties kunnen snel worden doorgevoerd.

Coface – voorbeeld uit de B2B-informatie

Sinds 1946 helpt Coface haar meer dan 120.000 klanten met credit ratings, B2B-informatie. Coface heeft een omzet van € 1,7 miljard, omvat 190 bedrijven en is actief in 97 landen met een staf van 6800 medewerkers. Met EasyNumber heeft Coface een online dienst die toegang geeft tot de gegevens van 60 miljoen bedrijven wereldwijd, groeiend tot 100 miljoen.

De uitdagingen voor Coface voor hun online dienstverlening waren als volgt:

- vergemakkelijk de toegang tot het groeiend aantal bedrijfsprofielen tot 100 miljoen;
- zorg dat het netwerk van gebruikers zich tot wereldschaal kan uitbreiden;
- indexeer en zoek ook met niet-latinse tekens als Chinees, Arabisch en Hebreeuws;
- bouw een open en uitbreidbaar platform dat toegang kan bieden tot allerlei nieuwe diensten, zoals data de-duplicatie en abonnementen op updates.

Na de selectie van Exalead CloudView is binnen 3 maanden het zoeken geïmplementeerd, compleet met integratie met bestaande toepassingen:

- unified access tot alle bedrijfsprofielen voor alle gebruikers;
- relevantere resultaten met een welhaast onbeperkt aantal criteria, fuzzy matching en navigatie op basis van facetten;
- zeer hoge indexeersnelheid, 60 miljoen profielen in 30 uur, en zoekperformance, gemiddeld 0,24 seconden per query;
- bijzonder snelle time-to-market van innovatieve diensten;
- lage TCO en open platform: 4 Linux servers, open API en gebruik van SOA.

Coface is bezig om meer bedrijfsprofielen vast te leggen, meer klanten te krijgen en nieuwe diensten te ontwikkelen.

Conclusie

Zoektechnologie is een krachtig platform, inclusief ETL en schaalbaarheid, om grote datasets voor veel gebruikers met uitgebreide functionaliteit te ontsluiten. Search-Based Applications vormen een flexibele manier om daarop veel specifieke BI-toepassingen te bouwen en het meeste uit uw investeringen in BI te halen en waarde voor uw organisatie te creëren. Met CloudView heeft Exalead als geen ander haar software toegesneden op deze aanpak.

Ronald Baan is directeur van EMID Consult.



BI-initiatieven mislukken nog veel vaker dan andere projecten. De BI-initiatieven moeten van de harde en de zachte kant komen. En als de harde kant van BI al praat met de zachte, spreken ze niet elkaars taal. Het boek BI-ware is een boek voor ICT'ers en voor managers en vertelt in gewoon Nederlands wat er allemaal fout kan gaan en wat daaraan kan worden gedaan. BI-ware bevat een bundeling van artikelen van Karien Verhagen en is een nieuwe uitgave in de reeks van DB/M Essays. De artikelen zijn gepubliceerd in de periode 2002 – 2006.

Wilt u weten hoe u Business Intelligence kunt laten slagen? Dan kunt u niet zonder deze uitgave! Ga snel naar www.array.nl en bestel BI-ware!

Deze uitgave is mogelijk gemaakt door: **Getronics** **PinkRocade**

DB/M

Array PUBLICATIONS