

Navigation Engine van Endeca kan geduchte concurrent worden

# BI uit een ander vat

Paul van der Linden

**Endeca is een relatieve nieuwkomer in de BI-markt (start in 1999) en ontleent haar naam aan het Duitse 'ontdecken', wat zoveel als ontdekken betekent. Dat is ook precies wat Endeca wil aanbieden: een product waarmee bedrijven zowel zichzelf als de buitenwereld kunnen ontdekken.**

Met het gelijknamige basisproduct, dat in 2001 in de markt werd gezet, is Endeca bezig om die buitenwereld stukje bij beetje te veroveren. Endeca is opgezet door mensen die niet afkomstig zijn uit de BI-wereld, maar die ervaring met ongestructureerde data en search engines meebrachten (onder andere Inktomi). Het resultaat is een BI-tool dat zich wezenlijk weet te onderscheiden van de rest van het BI-veld.

Endeca heeft ervoor gekozen om haar zoektechnologie initieel in de e-commerce sector te introduceren en wel onder de naam Endeca InFront. Achtien maanden geleden is ook het gebied van de enterprise search betreden, waarin producten als Verity, Autonomy en Convera traditioneel de scepter zwaaien. In deze markt heet het product Endeca ProFind. Kijkend naar de code van InFront en ProFind kan volgens Jeff Boehm (Director Product Management) geconstateerd worden dat het product voor 90 procent identiek is. De laatste 10 procent betreft specifieke toevoegingen. In het geval van de Navigation Engine, bedoeld voor de analytische wereld, gaat het dan bijvoorbeeld om de ondersteuning van verschillende grafieksoorten en de geboden visualisatie.

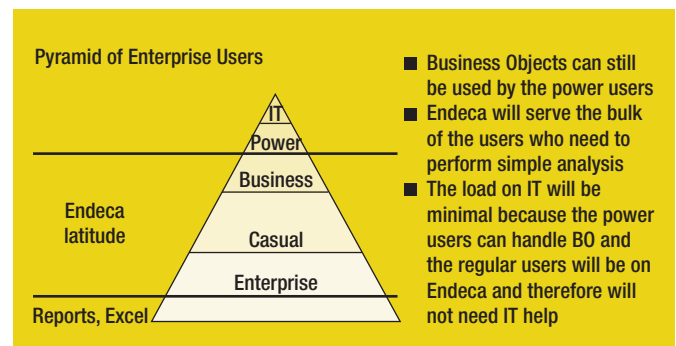
## Anders

De kaarten lijken niet slecht te liggen. Endeca voegt een nieuwe dimensie toe aan BI. Organisaties die gewend zijn aan Business Objects of Cognos PowerPlay zullen in Endeca een tool ontdekken die een meer intuïtieve manier van data-analyse mogelijk maakt. De wijze waarop Endeca de informatie aanbiedt en ordent lijkt op de werkwijze van een browser. Door het invoeren van een zoekterm verschijnt een lijst van resultaten. Echter, op basis van de data in de onderliggende bestanden maakt Endeca automatisch een dimensie aan. Deze dimensies zijn hyperlinks/selectievelden waarmee een verder inzoomen mogelijk wordt. Selecteren kan daarbij zowel op gestructureerde velden (bijvoorbeeld tekst) als op een afbeelding behorend bij ongestructureerde data (zoals bijvoorbeeld een Word-document of een PowerPoint-presentatie).

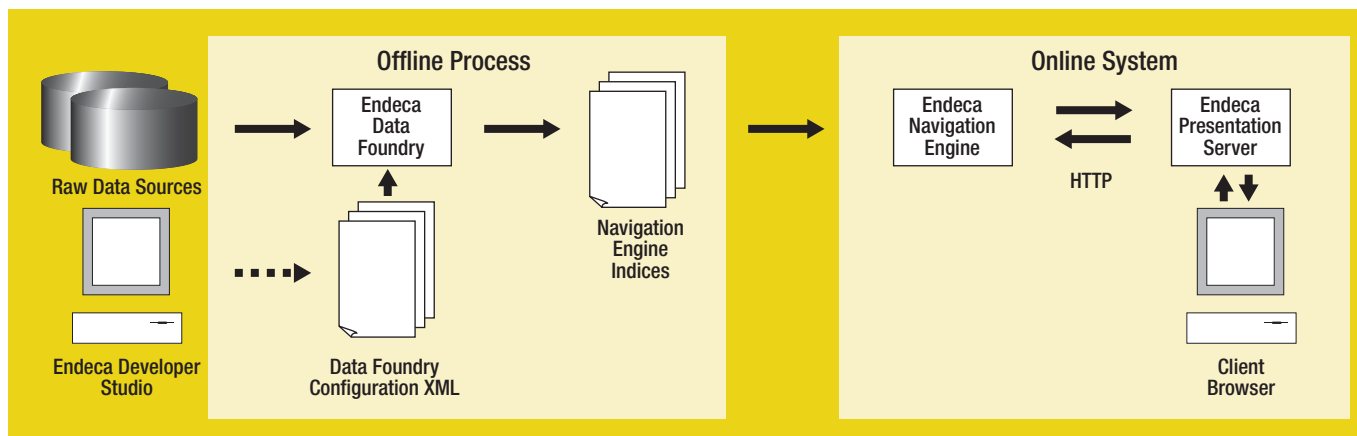
De presentatie van Endeca is overigens volledig programmeerbaar. Het kan er dus uitzien als een browser-scherm, maar elke andere vorm is ook mogelijk. Wie wil drillen (up, down en across) kan hiervoor zorgen. Omlaag en omhoog drillen is natuurlijk al mogelijk langs de automatisch gegenereerde hiërarchieën. Drill across zal geprogrammeerd moeten worden.

Behalve de intuïtieve manier om met informatie om te gaan is ook de snelheid waarmee data worden aangeboden indrukwekkend. Ook hier maakt Endeca gebruik van haar oorsprong in de internet-wereld waar het omgaan met grote hoeveelheden data en caching aan de orde van de dag is. Hoe werkt het? Endeca ontvangt haar gegevens uit operationele bronnen of uit een reeds aanwezig datawarehouse. Met behulp van de Foundry en Endeca Studio wordt deze informatie voorbereid en klaargezet om opgenomen te worden in een file-systeem. Dit file-systeem is zwaar geïndexeerd, hetgeen verklaart hoe een snelle responstijd kan worden behaald.

Endeca kan de operationele gegevens zowel via een push- als via een pull-mechanisme ontvangen. Daarnaast gevraagd geeft Vinay Mohta, product manager bij Endeca, aan dat er geen best practice bestaat. In de praktijk wordt volgens hem echter vaak gekozen voor een pull-mechanisme. Dit is conform de normale gang van zaken in een datawarehouse-omgeving. Endeca benadrukt echter dat het geen datawarehouse is. Als een organisatie reeds over een datawarehouse beschikt zal Endeca hier dankbaar gebruik van maken. Is dat datawarehouse er niet, dan is Endeca zelf in staat om haar gegevens uit de operationele omgeving te betrekken en



**Abbeelding I:** De coëxistentie met Business Objects.



**Afbeelding 2:** Overzicht van de Endeca basis-architectuur.

vervolgens te bewerken, zodat ze aan gebruikers ter beschikking gesteld kunnen worden. Maar ook in dit geval is er geen sprake van een datawarehouse en zijn de verzamelde gegevens niet bedoeld voor andere applicaties dan Endeca zelf.

Endeca ziet een aantal gaten in de BI-markt waarin het zich nu probeert te begeven. Het eerste gat is dat tussen OLAP-tools en reporting tools. Inderdaad ziet de praktijk er zo uit dat een keuze voor het ene aspect betekent dat het andere aspect er bekaaid afkomt. Het tweede gat dat men wil vullen is dat tussen de power users (een beperkte groep gebruikers) en de gebruikers van statische rapportage-tools. Wat Endeca hier wil bieden is een interactieve tool die door iedereen gebruikt kan worden – zonder dat men daarvoor een cursus hoeft te volgen. Zo'n eenvoudig bruikbare tool houdt ook in dat gebruikers zonder (veel) ondersteuning door IT aan de slag kunnen, waardoor de IT-backlog kan worden teruggedrongen.

## Architectuur

Hoe ziet de process flow eruit? Vanuit de verschillende aanleverende bronnen worden data aangeleverd die bestemd zijn voor de Foundry. De aangeleverde data kunnen flat files zijn, maar ook bijvoorbeeld relationele of XML-data. Met behulp van Endeca Studio worden deze aangeleverde data bewerkt, zodat ze geschikt zijn om opgenomen te worden in de Foundry. Studio kan gezien worden als een lichtgewicht ETL-tool, maar dan alleen bedoeld om data klaar te stomen ten behoeve van de Foundry. Op basis van de getransformeerde data bouwt de Foundry vervolgens hiërarchische modellen.

Uitgaande van deze modellen kunnen nu dimensies worden gedefinieerd of verder worden uitgewerkt. Ook de zoekregels worden hier beheerd. Het eindresultaat levert input voor de Indexer. Dit is het onderdeel waar de indexen (eigen indexformaat) worden aangemaakt en onderhouden. Tenslotte is het de Navigation Engine die het 'ontdekken' van informatie mogelijk maakt door middel van de gegenereerde indexen.

Opvallend is dat Endeca in de Foundry een aanpak hanteert die afwijkt van de aanpak die kenmerkend is voor een rdbms-

omgeving. Normaal worden daar eerst de tabellen gedefinieerd om vervolgens de data te laden. De structuur van de tabellen bepaalt daarbij de mogelijke zoekpaden. Wat de Foundry doet is op basis van de data (inhoud) een model bouwen. Uitgangspunt hierbij is dat de data aangeven welke zoekpaden mogelijk zijn. Het resultaat hiervan is dat alle mogelijke manieren om de data te doorzoeken nu mogelijk zijn. Uiteraard kunnen onlogische manieren ook weer eruit worden gehaald.

Wie bekend is met kennis-management en producten van bijvoorbeeld Collexis en Infolution (het vroegere Tarchon) zal deze aanpak herkennen. Ook hier wordt op basis van data-inhoud een navigeerbare structuur gegenereerd. Taxonomieën spelen hierin ook een belangrijke rol. Boehm geeft aan dat het hier om activiteiten gaat die Endeca heeft neergelegd bij een gespecialiseerde partner, Stratify. Stratify is eveneens in 1999 ontstaan en biedt software voor het managen van ongestructureerde data. De snelheid die Endeca weet te bieden wordt verkregen door de gegenereerde indexen en data in het geheugen te laden. Dat betekent dat voldoende RAM-geheugen aanwezig moet zijn. Algoritmen zorgen ervoor dat de indexen over meerdere processoren gedistribueerd kunnen worden.

## Relatie met BO

Begin jaren negentig was Jeff Boehm de product manager van Business Objects. Hij vindt het nog steeds een van de beste BI-suites in de markt, maar is van mening dat BO (overigens net als Cognos) zich voornamelijk richt op de power users en business-analisten. Dat is ook de reden dat Endeca BO niet ziet als een concurrent, maar eerder als een aanvullend product. Volgens Boehm is BO een high end BI-tool en is Endeca bedoeld voor de grote meerderheid van de BI-gebruikers. Dat Endeca veel intuïtiever is valt inderdaad niet te ontkennen. Om BO vervolgens een high end BI-tool te noemen is echter op zijn minst discutabel. Boehm constateert dat Endeca en BO een heel andere positionering hebben. Net zoals veel andere BI-leveranciers heeft BO ervoor gekozen om een volledig productenpakket aan te bieden van ETL-tools tot en met de verschillende front-end tools. Endeca Navigation Engine is in wezen een front-end tool.

De praktijk is echter weerbarstig. Want in selectietrajecten moet de Navigation Engine het vaak opnemen tegen Cognos en ... BO.

## TCO

Eén van de argumenten die Endeca gebruikt is dat de Navigation Engine een lagere TCO kent. De redenering hierachter is dat als meer mensen sneller de gezochte informatie vinden, dit minder kosten met zich meebrengt. De gekozen technologie maakt het mogelijk om op zwaardere hardware te besparen. Door een intuïtiever interface wordt een kleiner beroep gedaan op de IT-afdeling. Sneller en beter kunnen zoeken zou ook moeten leiden tot een hogere conversie, meer orders et cetera. Dit klinkt allemaal logisch, maar desgevraagd blijkt dat harde cijfers om het een en ander te onderbouwen nog ontbreken. Gezien het feit dat de Navigation Engine pas in het vierde kwartaal van dit jaar wereldwijd gelanceerd zal worden, mag niets anders verwacht worden. Zodra die cijfers er wel zijn heeft Endeca hiermee een goed commercieel verhaal. Nu blijft het nog (?) even steken in marketing-retoriek.

Tot de ondersteunde platformen behoren Windows, Linux en Sun Solaris (Intel/Sparc). Endeca kan naadloos aansluiten op verschillende software-pakketten zoals Interwoven en Documentum (beide document-management software), maar ook op ERP-, SCM- en LCM-systemen. In de e-commerce product search-markt maken onder andere Barnes & Noble, 1-800-Flowers, Tesco en WalMart gebruik van Endeca software.

Over de precieze kosten voor de Navigation Engine kan Endeca op dit moment alleen in globale termen iets melden aangezien het definitieve woord daarover nog niet gevallen is. Op basis van de ervaringen met ProFind en InFront zal de prijs van de Navigation Engine tussen \$ 300.000 en \$ 400.000,- liggen. Deze aanschafprijs heeft betrekking op een periode van drie jaar. Dit betekent

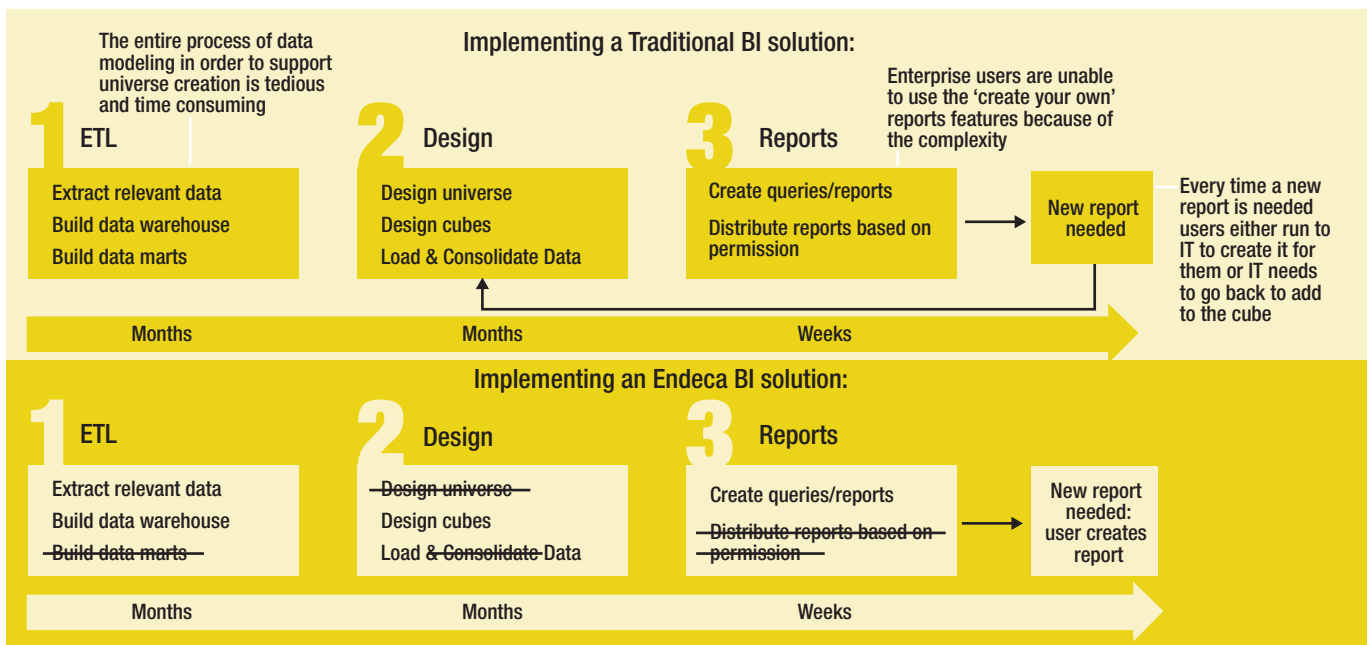
overigens niet dat bedrijven die tot aanschaf overgaan dit bedrag direct op tafel moeten leggen. De betaling kan gespreid plaatsvinden. Andere prijsconstructies zijn ook bespreekbaar. In de praktijk blijkt 80 procent van de bedrijven de voorkeur te geven aan de drie-jaarsperiode.

## Conclusie

Is de Endeca Navigation Engine het waard om ontdekt te worden? Het antwoord luidt zonder meer: ja. Vooral het intuïtieve karakter ervan in combinatie met de geboden snelheid, vormen een ijzersterke combinatie die gebruikers zeer zal aanspreken.

In organisaties die beschikken over massa's data en veel BI-gebruikers is Endeca zeker een product om naar te kijken. Als daar nog bijkomt dat BI voor een aanzienlijk deel ook betrekking heeft op ongestructureerde data, vormt het een moeilijk te kloppen product. In dat geval (veel data, veel gebruikers, combinatie van gestructureerde en ongestructureerde data) zal de aanschafprijs ook geen struikelblok blijken te zijn. De vooruitzichten voor Endeca zijn dus zonder meer goed. Echter, de Navigation Engine wordt pas in het vierde kwartaal van dit jaar gelanceerd. Er wordt op dit moment nog steeds aan gebouwd. Welke specifieke BI-functionaliteiten uiteindelijk worden geboden moet dan ook worden afgewacht. Dit zal in ieder geval gelijk moeten zijn aan hetgeen BO en Cognos op dit gebied bieden. Zo niet, dan is de Navigation Engine niets anders dan een dure zoekmachine die slechts basis-analysemogelijkheden biedt. Maar als men erin slaagt om op analysegebied tenminste dezelfde functionaliteit te bieden als de traditionele BI-spelers dan hebben deze er een geduchte concurrent bij. Alle intenties van coëxistentie ten spijt.

**Paul van der Linden** (Paul.PFH.vanderLinden@AtosOrigin.com) is senior consultant Data Warehousing/BI bij Atos Origin.



Afbeelding 3: Een traditionele BI-oplossing versus een Endeca BI-oplossing.