

De speld in de explosief groeiende hooiberg

Een nieuwe generatie BI-tool: CXAIR

Edwin van Asch

Steeds meer informatie is vandaag de dag elektronisch beschikbaar, waardoor ondernemingen ook steeds grotere hoeveelheden data verzamelen. Dat ondernemingen meer data opslaan, betekent niet automatisch dat uit die data ook meer informatie wordt gehaald. De hooiberg met data wordt almaar groter en de speld is steeds moeilijker te vinden. De relevante en direct bruikbare informatie zit steeds dieper verstopt in de explosief groeiende hoeveelheid data.

Deze groei brengt met zich mee dat de klassieke manier van omgaan met data en het verzamelen van informatie uit die data lang niet altijd meer het gewenste resultaat oplevert. Met Business Intelligence zoals we dat vandaag de dag kennen zijn de eindgebruikers afhankelijk van een complex IT-proces dat doorgaans bestaat uit een groot aantal stappen, zoals het maken van een extractie van de ruwe data uit de operationele systemen, het opschonen van de data en het transformeren naar een datawarehouse. Het uiteindelijke resultaat van dit proces bestaat doorgaans uit standaardrapportages die een deel van de standaard vragen beantwoorden maar lang niet alle ... en daar wringt de schoen!

Standaard rapportages en antwoorden

Vanzelfsprekend dekken standaardrapportages een belangrijk deel af van onze behoefte aan informatie. Echter, wat zijn onze mogelijkheden als we net even buiten de gebaande paden zoeken naar antwoorden en nieuwe inzichten? Wat als mijn vraag per definitie niet past in een standaard rapportage omdat ik op zoek ben naar antwoorden op ad hoc vragen? Voorbeelden van bedrijfsonderdelen die met dit soort ad hoc vragen zitten zijn: marketing, actuaariaat, product en category management en risk control. Zij zoeken naar nieuwe inzichten, verrassende combinaties en afwijkende patronen en die worden niet gevonden in standaardrapportages.

Hoe beantwoordt u een vraag als: welke producten worden vaak gekocht in combinatie met welke andere producten; de zogenaamde shopping basket analyse? Dergelijke analyses leveren een schat aan informatie op over koopgedrag van consumenten. Echter, hoe analyseert u *zelf veilig* en *ad hoc* enkele honderden miljoenen records van de afgelopen jaren om hieruit de relevante en direct toepasbare informatie te vinden?

Nieuwe 'drivers' voor BI

De slagkracht van een onderneming hangt mede af van het vermogen om informatie te halen uit de steeds sneller toenemende hoeveelheid data, zowel binnen als buiten de onderneming, zowel gestructureerd als ongestructureerd. Nieuwe methodes moeten worden geadopteerd om deze schat aan informatie te benutten. Business Intelligence tools moeten eenvoudig en met een dosis *fun* kunnen worden gebruikt door de kenniswerker zelf en in het verlengde liggen van de gereedschappen waarmee zij gewend zijn dagelijks te werken. Daarbij moeten de BI-tools de kenniswerkers de mogelijkheid bieden om te allen tijde zelf veilig de enorme hoeveelheden data die hen ter beschikking staan, te analyseren. Business Intelligence moet niet uitsluitend een tool zijn om informatie aan te bieden langs de gebaande paden; het moet de kenniswerkers ook zelf veilig de mogelijkheid bieden om de antwoorden te vinden op de vragen die zij hebben ten aanzien van onderwerpen waarover zij beslissingen moeten nemen.

Een nieuwe generatie BI-tools

De nieuwe generatie Business Intelligence tools pakt de zaken geheel anders aan en laat de traditionele databasetechnieken achter zich. De nieuwe generatie tools maakt gebruik van zoekmachine technologie en dát wat vandaag de dag in ruime mate én tegen lage kosten aanwezig is om veel data op te slaan; schijfruimte! Zoekmachine technologie is bijzonder snel en tevens in staat om deze snelheid ook te garanderen op zeer grote data-volumes. Daarbij is vrijwel iedereen gewend om met een zoekmachine te werken. De nieuwe generatie tools biedt verder de mogelijkheid om direct te werken met data uit operationele systemen maar kan ook bestaande datawarehouses, data-marts of Data Vaults als bron benaderen. Ondernemingen zijn hun datawarehouse soms misschien teveel als een doel op zich gaan

beschouwen, terwijl het eigenlijk een middel is om tot een resultaat te komen. De nieuwe technologie biedt de mogelijkheid om versneld meer rendement uit een bestaand datawarehouse te halen en de kenniswerkers te voorzien van de mogelijkheid om zelf veilig met deze data te werken, ongeacht de omvang. De gebruikersinterface is browser gebaseerd en biedt de mogelijkheid om de informatie gemakkelijk te ontsluiten via smartphones en tablet pc's.

CXAIR

CXAIR is een nieuwe generatie BI-tool die in Engeland wordt ontwikkeld door Connexica, de productnaam staat voor 'ConneXica Ad hoc Interactive Reporting'. In Nederland wordt CXAIR op de markt gebracht door Systemation uit Gouda. Eindgebruikers krijgen met CXAIR toegang tot alle data, zonder van te voren te bepalen via welke paden de data doorzocht moeten worden. CXAIR is geschikt voor het zoeken in ongelimiteerde hoeveelheden records, zonder dat de zoeksnelheid hierdoor beïnvloed wordt en zonder dat het effect heeft op het operationele systeem. Elke willekeurige zoekopdracht levert binnen luttele seconden het gevraagde resultaat op. De interface van CXAIR is beschikbaar vanuit een webbrowser.

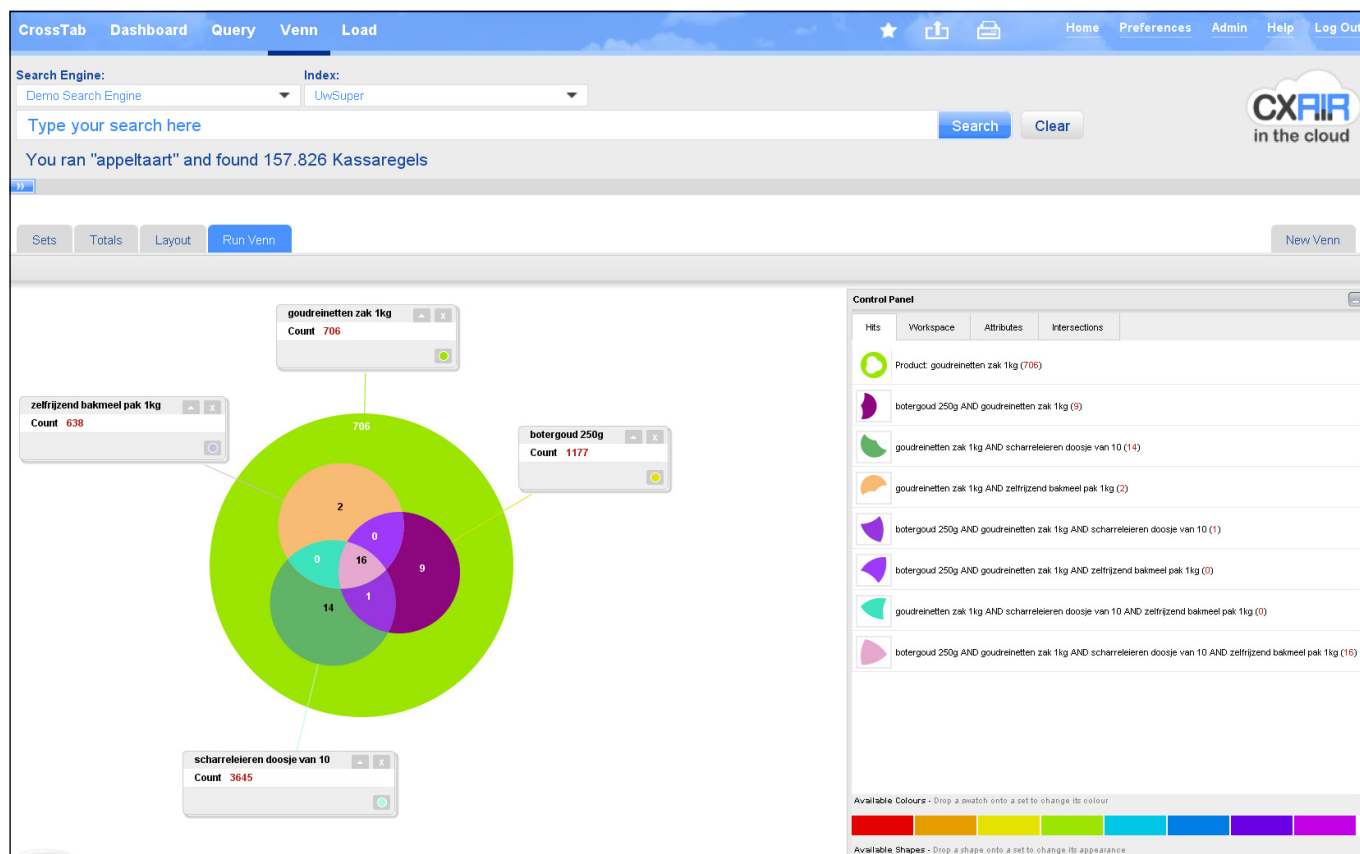
Het product CXAIR is ontstaan in de medische wereld waar behoefte was aan een oplossing die in staat is grote hoeveelheden onderzoeksresultaten te doorzoeken en te analyseren op bepaalde patronen. CXAIR is snel en gemakkelijk in het gebruik

omdat het gebaseerd is op zoekmachine technologie. De zoek-interface van CXAIR is vergelijkbaar met de bekende Google search bar. De indexeringstechniek is geschikt voor het rechtstreeks indexeren van transactionele databronnen, data-marts, Data Vaults en datawarehouses.

Een ander belangrijk aspect van CXAIR is het gebruik van harddiskruimte als opslag, deze is in tegenstelling tot het interne computergeheugen, waar in-memory databases gebruik van maken, in ruime mate en zeer goedkoop aan te schaffen.

De nieuwe generatie tools biedt de mogelijkheid om direct te werken met data uit operationele systemen

Het gebruik van CXAIR verloopt in vier eenvoudige stappen. *Step 1: Indexering.* In de eerste stap maakt CXAIR van bestaande databronnen een index zodat deze doorzocht kunnen worden. Dit indexeringsproces is te vergelijken met het indexeren van een website door een internet zoekmachine, maar in plaats van een website wordt door CXAIR een database doorzocht. Een index in CXAIR is een hoeveelheid data die bestaat uit rijen en kolommen. De zogenaamde referentietabellen die we kennen



Afbeelding 1: Venn-diagram.

uit een database zijn in de data als kolommen opgenomen. Een index kan worden samengesteld uit meerdere bronnen. Zowel identieke als gelijksoortige bronnen kunnen apart worden geïndexeerd en in CXAIR worden samengevoegd tot één index. Een voorbeeld van gelijksoortige bronnen is een supermarkt met meerdere merken kassasystemen, waarbij de kassaregels van de diverse kassa's, ondanks dat ze niet helemaal gelijk zijn, toch worden samengevoegd tot één index.

Stap 2: Zoeken. Na het indexeren kan een index in CXAIR worden doorzocht door het invoeren van zoektermen in een search bar. Willekeurige zoekvragen leveren ongeacht de grootte van de dataset altijd binnen luttele seconden het gevraagde resultaat op. Zoeken binnen teksten kan door het invullen van woorden, of delen van woorden met gebruik van wildcards. CXAIR gebruikt voor het zoeken een zogenaamde 'full text search', wat wil zeggen dat woorden op elke positie in de data worden gevonden. Voor de meer complexe zoekacties heeft CXAIR filters. Filters worden bijvoorbeeld gebruikt voor het zoeken in datumreeksen of het gebruik van 'van tot'-waardes in numerieke attributen. Voorbeelden hiervan zijn het zoeken van alle verkochte artikelen, in een bepaalde periode of met een prijs tussen de 2 en 5 euro.

Stap 3: Analyse met crosstabs en venn-diagrammen. Het resultaat van een zoekactie kan vervolgens verder worden geanalyseerd in de vorm van een crosstab, hierin worden meerdere grootheden tegen elkaar uitgezet. Een voorbeeld van een crosstab betreft de verkopen van bepaalde artikelen, per regio, per filiaal en per week. De werking van een crosstab is vergelijkbaar met een draaitabel in een spreadsheet. Aan een crosstab kunnen grafieken worden toegevoegd. De eindgebruiker kan de resultaten van een analyse bewaren voor hergebruik of om te delen met anderen. Naast crosstabs biedt CXAIR de mogelijkheid om verzamelingen grafisch te analyseren met behulp van venn-diagrammen. Een venn is een verzameling van gegevens uit een index met een bepaalde eigenschap die wordt weergegeven als een grafische vorm, bijvoorbeeld als een cirkel. Een voorbeeld van een venn is: "de verzameling van de verkopen van een bepaald product in een bepaalde tijdsperiode". Door meerdere van deze venns te combineren in een venn-diagram worden de relaties tussen de verschillende verzamelingen weergegeven. Een praktijkvoorbeeld van het gebruik van venn-diagrammen is het uitvoeren van de eerder genoemde 'shopping basket analyse'. Deze analyse is in de retailmarkt belangrijk om te bepalen wat de klant in één aankoop heeft aangeschaft. Door de venns van verschillende producten interactief over elkaar heen te slepen ontstaan direct doorsneden. Deze doorsneden geven, in het geval van de kassaregels in de supermarkt, aan hoeveel klanten bepaalde producten tegelijk hebben gekocht. Met andere woorden, welke producten staan samen op een kassabon? Afbeelding 1 toont een voorbeeld van een venn-diagram met vier producten die samen de basis vormen voor het bakken van

"Veel ondernemingen hebben te maken met een vloed aan data met inefficiënte software middelen en processen. Het kost gemiddeld per kenniswerker 18.000 dollar per jaar om informatie te vinden, wat resulteert in een groot verlies in productiviteit." – Bill Gates.

appeltaart. In het midden van dit venn-diagram is een gebied te vinden met daarin 16 klanten die de 4 producten voor het maken van een appeltaart; boter, meel, eieren en appels, tegelijk hebben gekocht. Deze selectie van 16 klanten kan vervolgens weer dienen voor verdere analyse. Wellicht wilt u vervolgens weten op welke dagen, in welke regio's en in welke winkels deze combinatie werd gekocht.

Stap 4: Presentatie met rapporten en dashboards. Wanneer u uw analyse gedaan heeft en uw heeft informatie gevonden, dan wilt u dit ook delen met collega's. CXAIR biedt u de vrije hand om zelf te bepalen hoe u dit wilt doen. U kunt een rapport maken en dit delen met andere gebruikers of het exporteren naar een ander formaat zoals bijvoorbeeld een Microsoft Office-formaat als Excel. Een andere manier is om informatie op te nemen in een dashboard. Een dashboard is een manier in CXAIR om data van verschillende crosstabs, grafieken of venn-diagrammen te verzamelen en als geheel te presenteren. Verder is het mogelijk in een dashboard ook data te tonen in de vorm een kaart. CXAIR ondersteunt 'Google maps' voor het weergeven van informatie op een landkaart, of zelfgetekende kaarten om bijvoorbeeld de aankopen van een supermarkt op een winkelindeeling te tonen.

Conclusie

CXAIR geeft eindgebruikers de beschikking over een herkenbaar stuk gereedschap om zelf veilig, ad hoc en interactief data te doorzoeken, analyseren en presenteren. Door het gebruik van zoekmachine technologie kunnen met CXAIR zeer snel grote hoeveelheden data worden doorzocht. Door de search bar is CXAIR eenvoudig in het gebruik en heeft het een lage leercurve. Verder is met name de manier waarop je interactief met venn-diagrammen verzamelingen van data kunt analyseren, uniek. De resultaten van een analyse (crosstab of venn-diagram) kan de eindgebruiker zelf presenteren in een rapport of een dashboard dat kan worden bewaard voor hergebruik of gedeeld met anderen. CXAIR is een complete BI-tool, die ook toegevoegde waarde biedt aan bestaande Business Intelligence projecten omdat deze helpt meer rendement te halen uit de reeds gedane investeringen.

U kunt dus blijven zoeken naar de speld door het hooi sprietje voor sprietje te verplaatsen tot u de speld in handen heeft. Echter, aangezien de hooiberg sneller groeit dan u kunt verplaatsen, is een andere methode het overwegen waard!

Edwin van Asch is Solution Consultant bij Systemation.