



Tools voor alternatieven self-service BI en workgroup BI

Ad hoc analyse & reporting reloaded

Wouter van Aerle

In het gebruik van Business Intelligence is ad hoc analyse & reporting steeds vaker als toepassingsvorm te herkennen. Strategische analyses, het beantwoorden van Kamervragen en afwijkende productevaluaties zijn praktijkvoorbeelden die passen in de definitie van ad hoc reporting, die eerder in DB/M is gepresenteerd.

Ontwikkelingen in de markt bevestigen dit beeld. Het afgelopen jaar zijn nieuwe producten op de markt gekomen en hebben bestaande leveranciers hun portfolio uitgebreid met nieuwe functionaliteit die expliciet op dit type gebruik is gericht. Daarnaast duiken in de vakliteratuur alternatieve typering op die lijken op de ad hoc definitie. Zoals we in dit artikel zullen laten zien gaat het in alle gevallen om hetzelfde. Genoeg redenen dus om ad hoc reporting in navolging van het eerste artikel¹ verder onder de loep te nemen.

Ad hoc reporting

Om het geheugen op te frissen zetten we de belangrijkste karakteristieken nog even op een rij:

- Er is sprake van een volledige end-to-end BI-ontwikkelcyclus: van data-integratie tot en met het opstellen en distribueren van rapportages;
- De eindgebruiker is degene die dit proces uitvoert, niet de IT-afdeling of een BI-ontwikkelaar;
- Er wordt een mix van gegevensbronnen gebruikt: datawarehouse data en operationele data, intern en extern;
- Op enig moment in het proces brengt de eindgebruiker specifieke kennis en ervaring in om een processtap uit te kunnen voeren, bijvoorbeeld bij het combineren van interne en externe gegevens of het maken van een specifieke analyse;
- Het hele proces heeft een iteratief karakter: de processtappen worden meerdere keren doorlopen.

Daarnaast is een aantal redenen genoemd waarom ad hoc reporting wordt gebruikt. Deze redenen hadden te maken met het beantwoorden van eenmalige of laagfrequente (management) vragen, prototypen maken van nieuwe rapportages, de noodzaak van een korte time-to-market (wanneer op korte termijn een businessvraag beantwoord moet worden of vooruitlopend op gestructureerde nieuwbouw) en als oplossing voor kleinere orga-

nisaties die niet of beperkt kunnen beschikken over een volwaardige IT-afdeling of BICC. We kijken naar twee gerelateerde typering: self-service BI en workgroup BI.

Self-service BI

Wayne Eckerson, research director van The Datawarehouse Institute (TDWI), geeft de volgende omschrijving van self-service BI: "Business users create their own custom reports using end-user oriented query and reporting tools running against the data warehouse".² Die business users worden onderverdeeld in 'information consumers' of 'casual users' en 'information producers' of 'power users'. Uit de aanvullende TDWI Checklist Report³ blijkt dat met casual users eindgebruikers worden bedoeld die een beperkt aantal vrijheidsgraden kennen bij het raadplegen van rapportages: zij maken geen rapporten maar gebruiken standaard rapportages waarbinnen ze de mogelijkheid hebben om door gegevens heen te navigeren door bijvoorbeeld *pivoting*, *drilling* en *prompting*. De power users "... are data savvy individuals who are comfortable accessing, analyzing and publishing data on a regular basis. [...] I call this type of BI self-service ad hoc report creation". Hier ligt duidelijk een sterke relatie met onze definitie van ad hoc reporting. We zoomen nog wat verder in op de overeenkomsten en verschillen:

- De noodzaak dat ad hoc reporting moet passen binnen de taken, verantwoordelijkheden en bevoegdheden van de medewerker is een duidelijke overeenkomst;
- De 'super user' is een subvariant waarbij "technically savvy business users take responsibility for creating ad hoc reports on behalf of colleagues [...] super users are an extension of a BI Competency Centre". Deze verantwoordelijkheid is te projecteren op het beoogd gebruik om prototypen van rapporten te maken en voor initiële rapportontwikkeling te zorgen (het time-to-market argument);

- Het integreren van gegevens krijgt daarentegen niet een expliciete plaats binnen de self service-typing. Aangenomen wordt dat de gegevens beschikbaar zijn in een datawarehouse. Onze definitie van ad hoc reporting is breder: het verzamelen, integreren en (voor analyse en reporting) voorbereiden van gegevens behoort uitdrukkelijk tot de scope. Bevinding 7 uit het Checklist Report – ‘Liberate power users’ – pleit overigens wel voor meer mogelijkheden voor power users om ook externe data te benaderen maar werkt dit niet verder uit;
- Power users zouden alleen rapporten voor zichzelf mogen maken, niet voor het grote publiek. Wanneer een zelfgemaakt rapport generiek beschikbaar gesteld moet worden, dient dit te worden goedgekeurd door een ‘report governance committee’ en in beheer te worden genomen door het BI-team. Zo’n werkwijze klinkt logisch maar moet wel afhankelijk zijn van de reden waarom een rapportage wordt opgesteld. Als ad hoc reporting is gebruikt voor prototyping of om de time-to-market te verkorten is een governance proces op zijn plaats, maar in geval van een eenmalige of laagfrequente vraag is dat overbodig.

Workgroup BI

In een meer recente publicatie van TDWI wordt de trend gesignaleerd van toenemende “explicit end-user oriented offerings”.⁴ Deze propositie wordt aangeduid met de term *workgroup BI*.

Producten die in dit kader worden genoemd zijn onder andere Lyzasoft, QlikView en Gemini van Microsoft.

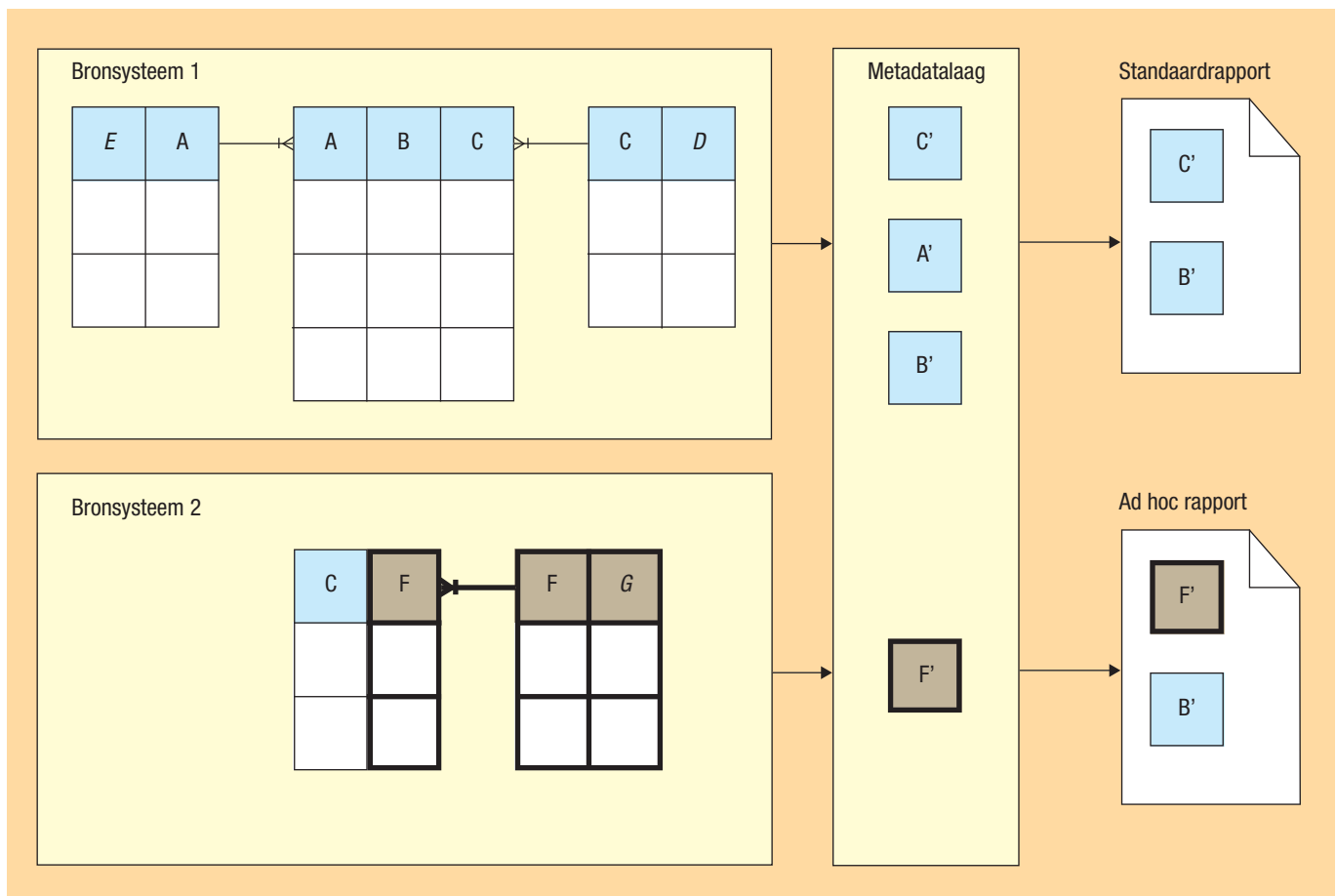
Workgroup BI verschilt feitelijk nauwelijks van self-service BI. Het verschil zit meer in wat er beschreven wordt: workgroup BI typeert de tooling, terwijl self-service BI meer het voortbrengingsproces van informatie typeert.

De beperkingen van data-integratie worden ook in het workgroup BI-artikel benadrukt: “Workgroup BI offerings tend to be sold with a caveat: namely, that IT must first perform the requisite background data integration”. Die noodzaak komt echter grotendeels voort uit de beperkte functionaliteit voor data-integratie van de tools: “In the case of QlikTech and Lyza, data integration (DI) [...] is primitive, at best taking the form of generic ODBC or JDBC connectivity”.

Beide artikelen benadrukken daarnaast het risico van *stove-piping*: de soms onbeheersbare groei van door gebruikers zelf ontwikkelde en beheerde rapportages. Ironisch genoeg is juist deze tendens – ook wel aangeduid met *spreadmarting* – in het verleden vaak aanleiding geweest voor het beperken van self-service mogelijkheden.

Samengevat zijn de belangrijkste bevindingen als volgt:

- Voor zover het power users betreft komt de definitie van self-service BI in hoge mate overeen met die van ad hoc reporting en de omschrijving van workgroup BI;



Afbeelding 1: Hergebruik.

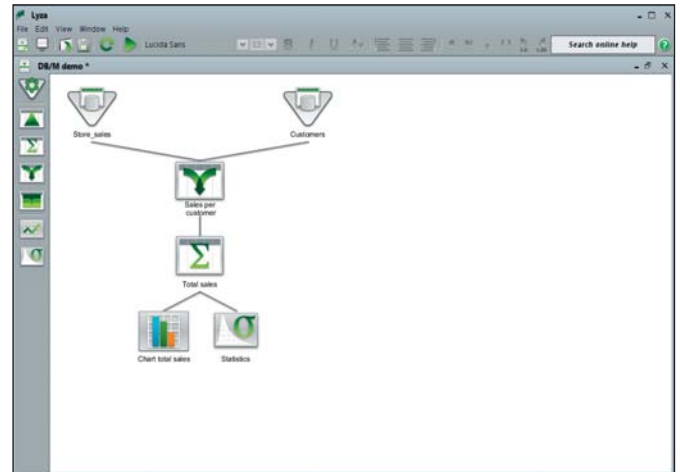
- Self-service BI moet passen binnen het functieprofiel van de medewerker: stel ad hoc reporting alleen beschikbaar aan gebruikers wanneer het past binnen hun taakomschrijving;
- Binnen het spectrum van BI-functionaliteiten is self-service BI één van de smaken, los van de vraag of de functionaliteit vanuit één enterprise BI-platform wordt ondersteund of door aparte tooling;
- Het ontwikkelen van standaardrapportages kan uitstekend door een self-service BI-proces worden gefaciliteerd. Voorwaarde is wel dat het om structurele en meer permanente rapportages gaat, waarbij de rapportages door middel van een vorm van centrale regie en beheer definitief worden gemaakt. De power user fungeert hierbij als verlengde van het BI-ontwikkelteam;
- Bij kleine organisaties is self-service BI de enige manier om rapportages te ontwikkelen. De schaalgrootte van dit soort organisaties is te klein voor centrale ontwikkeling door een apart BI-team. De power user is hier het BI-team;
- De kwalificatie 'ad hoc' in het eerste DB/M artikel¹ blijkt nu te beperkt. Het proces dat daar werd beschreven is eigenlijk een self-service manier van voortbrenging, waarbij het ad hoc karakter van de informatievraag vaak de reden is om deze voortbrengingswijze te gebruiken. Zoals overigens ook in het eerste artikel is beschreven bestaan er naast ad hoc-vragen ook andere redenen om self-service BI toe te passen. De aanduiding 'self-service' dekt de lading dus beter.

De rol van data-integratie blijkt tenslotte het grootste verschilpunt te zijn. Over het algemeen wordt aangenomen dat de data-integratie geregeld is voordat de power user zelf rapportages gaat ontwikkelen. In principe sluit self-service als concept data-integratie echter niet uit. In de praktijk wordt het alleen vaak buiten scope gehouden, omdat de tooling onvoldoende ondersteuning biedt zoals het workgroup BI-artikel ook aangeeft. Beheersmatig gezien zijn er ook argumenten voor: wanneer een power user nieuwe bronnen ontsluit, zal het BI-team bij inbeheername deze nieuwe data-integratie moeten overnemen. Dit kan betekenen dat alsnog een volledig ETL-proces moet worden ontwikkeld. In zo'n geval is het de vraag of het BI-team niet meteen van begin af aan de ontwikkeling voor haar rekening had moeten nemen. In het geval van een echte ad hoc-vraag (dus met een tijdelijk karakter) is data-integratie door de power user wel uitdrukkelijk gewenst. Denk hierbij aan een situatie waarin interne (datawarehouse) data eenmalig worden geconfronteerd met een extern bestand. Op het dilemma met data-integratie komen we bij de functionele eisen terug.

Functionele eisen

De inzichten uit de diverse publicaties over self-service BI worden nu gebruikt om een globale inventarisatie van functionele eisen op te stellen. Hierbij staat de power user rol centraal. Dit leidt tot het volgende overzicht:

1. Ondersteuning voor de hele BI-ontwikkelcyclus: van data-integratie tot en met rapportage- en distributiemogelijkheden.



Afbeelding 2: Screenshot van Lyza.

2. Gebruiksgemak: een self-service BI-tool wordt bediend door een eindgebruiker, niet door een software engineer. Dit stelt hogere eisen aan de user-interface.
3. Security en autorisatie: dit is vereist wanneer meerdere gebruikers aan dezelfde analyse werken, om te regelen wie over welke functionaliteit mag beschikken en om hergebruik van (centrale) metadata te faciliteren.
4. Performance en resource management: hoge performance is een must. Daarnaast zal een vorm van resource management nodig zijn wanneer van gemeenschappelijke infrastructuur gebruik gemaakt wordt. Voorkomen moet worden dat power users de performance en beschikbaarheid van reguliere BI-toepassingen kunnen beïnvloeden ('the query that dims the light').
5. Gebruikstatistieken en -monitoring: om regie te kunnen voeren op datgene wat ontwikkeld wordt, is een vorm van gebruiksmeting vereist.
6. Integratie met (de metadata van) het standaard BI-platform dat binnen de organisatie wordt gebruikt.
7. Audit trail: self-service BI stoelt niet op ontwerpen en documenteren; de gebruiker gaat gewoon aan de slag. Daarom is het van belang dat de tooling zelf inzicht verschaft in de stappen die de power user tijdens zijn analyse zet. Door automatisch een audit trail bij te houden – liefst gevisualiseerd – ontstaat inzicht en transparantie in de handel- en denkwijze van de analist. Dit helpt om achteraf te begrijpen hoe de resultaten tot stand zijn gekomen.

De eisen 1 tot en met 4 worden in meer detail beschreven in de achtste aanbeveling van het TDWI Checklist Report. De mate waarin tooling kan voldoen aan de zesde eis zal daarnaast onder andere het risico op stovepiping beïnvloeden en de mate waarin data-integratie effectief kan worden ondersteund. We lichten dit verder toe.

Vrijwel ieder BI-platform werkt met een metadata-laag. Simpel gezegd worden fysieke gegevensbronnen – doorgaans (attributen van) databasetabellen – vertaald in betekenisvolle bouwstenen,

de metadata-objecten. Met deze bouwstenen worden de gebruikerstoepassingen gemaakt (rapportages, dashboards, kubussen en dergelijke). In MicroStrategy heten deze bouwstenen de schema objects, in BO de universes en in SAS de information maps om een paar voorbeelden te noemen. Zeker daar waar een organisatie gestandaardiseerd heeft op één platform is het definiëren van de metadata-objecten cruciaal. Door gestructureerd deze mapping te maken, worden op een consistente, eenduidige en uniforme manier gegevens beschikbaar gemaakt. Zo'n aanpak voorkomt bijvoorbeeld dat hetzelfde gegeven twee keer als bouwsteen wordt opgenomen of dat bijvoorbeeld conformiteit van dimensies verloren gaat. Het aanmaken en beheren van de metadata gebeurt karakteristiek door een centrale autoriteit als een BICC.

In het geval van self-service is de gebruiker idealiter in staat deze bouwstenen te hergebruiken én nieuwe gegevensbronnen toe te voegen voor zijn specifieke analyse. Zie afbeelding 1 voor een vereenvoudigd voorbeeld.

Van bronsysteem 1 zijn gegevens in de metadatalaag gedefinieerd. A en C zijn de dimensies, B is een feittype en D en E (gecursiveerd) zijn attributen. Met C en B is een standaardrapport gemaakt. De power user wil nu op basis van bestaande gegevens/metadata-objecten een nieuw rapport maken waarin hij gegevens verder aggregereert op basis van gegevens uit bronsysteem 2: F is een aggregatie van C (bijvoorbeeld van product naar productgroep). De gebruiker moet dan in staat zijn om zelf F in de metadatalaag aan te maken – dit is immers de enige nieuwe informatie – en op basis van B (bestaande bouwsteen) en F een nieuw rapport te maken.

Belangrijk is dat hij dit moet kunnen doen door middel van een gebruikersinterface die voor hem als gebruiker is geoptimaliseerd (eis nr.2). Dus niet via de (thick) client waar de ontwikkelaars mee werken maar via een end-user interface. Goede securityvoorzieningen (eis nr. 3) moeten bovendien regelen dat de power user bestaande bouwstenen kan hergebruiken zonder dat hij ze kan aanpassen.

Tools

Ter afsluiting volgt een beknopte indruk van BI-leveranciers met een self-service propositie. Hierbij worden onder andere specifieke producten aangehaald om de verschillende functionele eisen te illustreren. Het bestek van dit artikel is echter te kort om een *long list* van producten te geven, alsmede aan ieder product evenveel aandacht te schenken.

Een van de eerste zaken die opvallen is een tweedeling in het aanbod. Enerzijds zijn er de leveranciers van bedrijfsbrede BI-platformen als SAS (Enterprise Guide), MicroStrategy (multi-source), SAP BO (BO Explorer) en Information Builders (WebFocus, InfoAssist) die als onderdeel van het platform self-service functionaliteit hebben opgenomen (tussen haakjes vermeld). Daarnaast is er een productaanbod van leveranciers die zich meer specifiek richten op het self-service gebruik als

Qliktech, LyzaSoft, Tableau en Spotfire. Gemini van Microsoft zit er tussenin, want het product is met name een end-user tool, terwijl Microsoft met SQL Server en Analysis Services ook een breed platform aanbiedt. Vooralsnog zijn deze producten nog niet met elkaar geïntegreerd.

Wat betreft user interface zijn interessante verschillen zichtbaar. Het meest opvallend is wellicht wel Microsoft dat met Gemini zwaar inzet op het gebruik van Excel als gebruikersinterface. Gemini is simpel gezegd een plug-in voor Excel met een eigen, gecomprimeerde in-memory column store en komt naar verwachting in 2010 op de markt. Bij QlikView bijvoorbeeld is de gebruikersinterface van begin af aan ontworpen met de eindgebruikers als primaire doelgroep in het achterhoofd. Andere leveranciers bieden Excel-plugins om via Excel de functionaliteit van het platform te kunnen gebruiken (onder andere SAS en MicroStrategy).

Lyzasoft heeft in haar product functionaliteit voor een visuele audit trail opgenomen, zie afbeelding 2. Duidelijk is te zien hoe de opeenvolgende bewerkingen die de gebruiker heeft uitgevoerd gevisualiseerd worden in stappen. Hierdoor ontstaat de benodigde transparantie; iets wat vaak compleet onzichtbaar wordt bij zwaar gebruik van Excel als self-service oplossing. Dergelijke functionaliteit wordt ook geboden door Enterprise Guide van SAS. In-memory opslag is daarnaast een techniek die door veel leveranciers aan producten is toegevoegd om hoge performance te garanderen.

Conclusie

Ad hoc reporting, workgroup BI en self-service BI zijn verwante typeringingen gebleken die in de basis op hetzelfde betrekking hebben. Onze voorkeur heeft uiteindelijk de aanduiding self-service BI. Duidelijk is geworden dat self-service BI als een alternatief voortbrengingsproces gezien moet worden dat voor specifieke doelen gebruikt wordt en in specifieke omstandigheden geschikt is. Dit proces wordt steeds beter ondersteund door diverse bestaande en nieuwe BI-tools. Data-integratie is bij de diverse tools een specifiek punt van aandacht.

Noten

1. *Ad hoc reporting & analyse onder de loep*, R. van Lun & W. van Aerle, DB/M nr. 5, augustus 2008.
2. *The myth of self-service BI*, W. Eckerson, TDWI, Vol. 24, november 2007.
3. *TDWI Checklist Report Self-service BI*, TDWI, juli 2009.
4. *Workgroup BI poised for a comeback*, S. Swoyer, TDWI, 25 maart 2009.

Wouter van Aerle

Ir. W. van Aerle (wouter.van.aerle@capgemini.com) is managing consultant in de service line Business Information Management bij Capgemini. Met dank aan Jos van Dongen.