

Succesvol implementeren van een enterprise datawarehouse

Vier belangrijke faal-factoren in kaart gebracht

Ad Stam

Binnen veel grote organisaties zijn projecten in uitvoering of in voorbereiding met als doel de Business Intelligence-voorziening te verbeteren door het implementeren van een organisatiebreed datawarehouse.

In de praktijk blijkt het echter heel moeilijk om dergelijke ambitieuze projecten tot een goed einde te brengen. Daarvoor is een viertal hoofdredenen aan te wijzen: geld (en dus governance); het ontwerp; erfenis; de data. Dit artikel beschrijft deze vier belangrijke faalfactoren.

Geld

Een oud gezegde luidt dat "wie de muziek betaalt het gespeelde wijsje bepaalt". Het geld voor IT-projecten wordt opgebracht door de business op basis van de toegevoegde waarde zoals die door de (toekomstige) gebruikers gepercipieerd wordt. Althans zo hoort het te zijn. In het geval van een datawarehouse is de toekomstige gebruikersgroep het management en de ondersteunende staf. De logica van een enterprise datawarehouse (*the single version of the truth* en "het optimaliseren van de gegevenstromen tussen N bronnen en M Business Intelligence-applicaties") is op zich zelf helder en eigenlijk onweerlegbaar. Maar dan doet zich een ander fenomeen voor en ook daarvoor is een gezegde. "Managers en kinderen hebben één ding gemeen, als ze iets willen dan willen ze het meteen". En een enterprise datawarehouse heb je nu eenmaal niet meteen. Daar komt bij dat managers een informatie-behoefte hebben die voor een belangrijk deel snel wisselt en onvoorspelbaar van aard is, terwijl hij tegelijkertijd eigenlijk 'nu' ingevuld moet worden. En dus is de verleiding groot om toch maar weer een *short cut* buiten het enterprise datawarehouse om te handhaven of bouwen.

Dus hoewel de manager het nut van een enterprise datawarehouse zal onderschrijven, zal de bereidheid om het vanuit *zijn* project te laten financieren niet echt groot zijn. Dat geldt ook voor *zijn* bereidheid om voor de invulling van zijn informatiebehoefte te wachten tot het datawarehouse 'zover is'.

De les die hieruit getrokken kan worden is dat de financiering

van het enterprise datawarehouse ontkoppeld moet worden van individuele Business Intelligence-projecten. Het is een soort 'basisinfrastructuur' die pas na het doen van een aanzienlijke drempelinvestering een rol kan spelen in de feitelijke informatievoorziening aan de klanten ervan. Overigens moet de invulling van het datawarehouse, dus welke gegevens uit welke systemen komen er wanneer in, wel degelijk bepaald worden door de prioriteit van de verschillende BI-projecten.

Ontwerp

Een tweede reden waarom enterprise datawarehouse-projecten moeilijk tot succes komen ligt in het interne ontwerp ervan. Eén van de belangrijkste, zo niet de allerbelangrijkste, gestelde eisen is het vermogen om de veranderingen in de business te volgen in het tempo dat door business gedictieerd wordt. Managers willen informatie en welke interne of externe veranderingen zich ook voordoen, zij willen dat hun informatievoorziening 'naadloos' doorgaat en een antwoord geeft op de nieuwe vragen. Sterker nog, zij willen antwoorden op vragen, die ze pas bedenken op het moment dat het antwoord op de vorige vraag op hun bureau ligt. Dit betekent dat als de datawarehouse-organisatie niet in staat is in het vereiste tempo te reageren, gebruikers afhaken en wederom of alsnog hun eigen informatievoorziening los van het datawarehouse zullen optuigen.

Het datawarehouse is in feite een logistieke dienstverlener die zorgt dat gegevens die ontstaan in een aantal bronsystemen in een aantal doelsystemen binnen de gegeven *time to market* als zinvolle informatie gebruikt kunnen worden. Hiervoor dient een logistiek proces ontworpen en ingericht te worden met verzamel-punten, transportmiddelen, ontkoppelpunten waarin (tijdelijke) opslag en bewerking plaatsvindt en distributiepunten. Een dergelijk logistiek proces wordt in afbeelding 1 conceptueel weer-gegeven.

Bij het ontwerpen van de processen, systemen en applicaties in dit proces dient – om de gewenste flexibiliteit en snelheid van veranderen te bereiken – rekening te worden gehouden met het toepassen van vier ontwerpprincipes: generiek in plaats van specifiek; herbruikbaar in plaats van éénmalig; ontkoppelen in plaats van integreren; federatief in plaats van centralistisch.

Generiek in plaats van specifiek.

Een datawarehouse is uiteindelijk een normale database, die bestaat uit een aantal fysieke relationele tabellen. Een dergelijk fysiek database-model is gebaseerd op een logisch database-model, waarin de relaties tussen de verschillende informatie-elementen zijn vastgelegd. Op zijn beurt is het logische database-model weer gebaseerd (althans dat hoort zo te zijn) op het informatie- of business-model. Dit informatiemodel is de manier waarop de gebruiker tegen zijn 'wereld' aankijkt. De meest ingrijpende wijzigingen in een enterprise datawarehouse-omgeving hebben veelal te maken met wijzigingen in het business-model en daaraan gekoppeld in het logische en het fysieke datamodel van het datawarehouse. Toevoegingen van dimensies of wijzigingen in de structuur van dimensies vragen veel onderhoud, niet alleen van het datamodel zelf, maar ook van de laadprocessen tussen bronsysteem en datawarehouse enerzijds en tussen het datawarehouse en datamarts anderzijds. Oplossingen worden gezocht in het gebruiken van standaard (generieke) bedrijfstakmodellen (bijvoorbeeld van IBM of Teradata) of in het gebruik van een datawarehouse, waarbij standaard de business content van het onderliggende ERP-systeem wordt ingeladen (bijvoorbeeld SAP of Peoplesoft). Echter in de wat grotere organisaties is standaard nooit standaard.

Bedrijfstakmodellen hebben aan de ene kant vaak teveel inhoud en moeten aan de andere kant uitgebreid of aangepast worden, al was het maar omdat elke zichzelf respecterende business zich zo uniek vindt dat er toch minstens het nodige 'aangepast' (gecustomized) moet worden. Het alternatief om zichzelf dan maar naar de standaardoplossing te organiseren komt vrijwel nooit voor, maar

zou – tenminste als de kosten van reorganisatie lager zijn dan de kosten van het aanpassen van de systemen – wel het meeste rendement uit de investering halen.

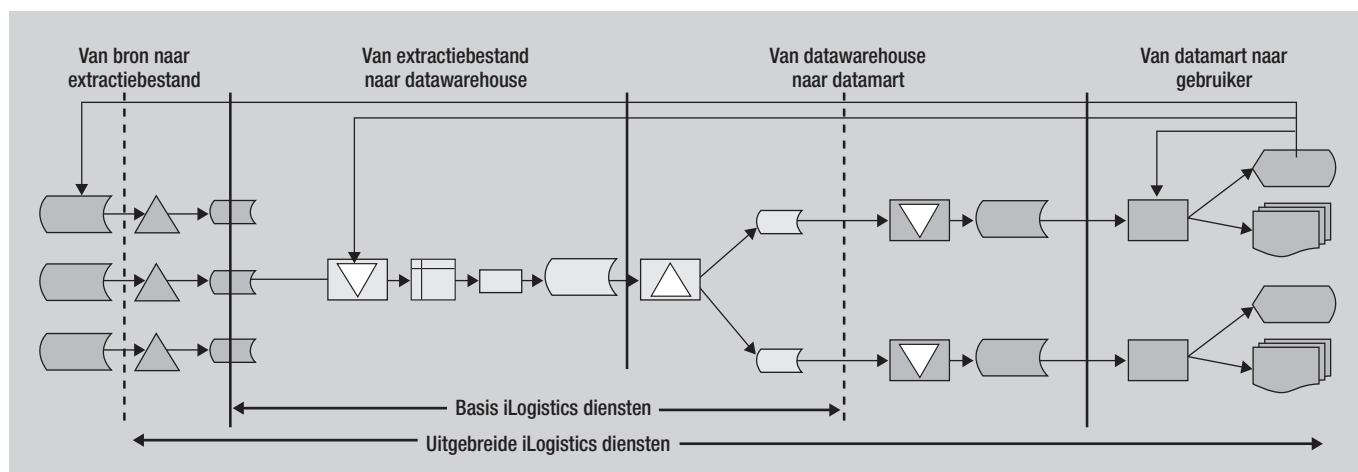
Standaard business content werkt (en bijna altijd ook maar als het onderliggende ERP-systeem niet aangepast is) voor dat ERP-systeem. Helaas is dat nooit het enige bronsysteem en zijn er dus wijzigingen nodig voor de andere bronnen. En al die wijzigingen worden uiteindelijk gewoon weer doorgevoerd in een 'normaal' datamodel. Het resultaat is dus dat je – afhankelijk van de mate van initiële dekking – hooguit een 'headstart' krijgt, maar dat daarna toch weer onderhoud in een 'klassiek' model uitgevoerd moet worden.

De omvang en ernst van problemen van fouten in brondata worden in eerste instantie vaak genegeerd

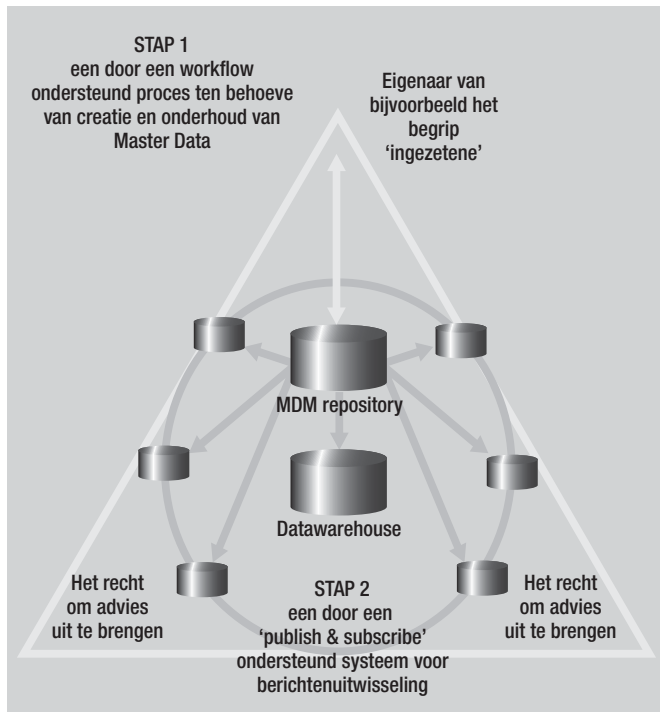
Maar gelukkig zijn er andere mogelijkheden. Zo is er bijvoorbeeld Datavault als datamodelleringsmethode. Datavault is een door een Amerikaans adviesbureau ontwikkelde methode waarin het logische datamodel voor het datawarehouse op een (semi)-generieke wijze wordt ondersteund. Verder is er Kalido, een product dat oorspronkelijk midden jaren negentig ontwikkeld is binnen Shell, dat een volledig generiek datamodel gebruikt, waarbij wijzigingen in het business-model alleen maar leiden tot mutaties in of toevoeging van nieuwe rijen in het *generieke* logische en fysieke datamodel. Andere producten (bijvoorbeeld van twee kleine Nederlandse start-up's namelijk Dynalytical en BIReady) genereren vanuit wijzigingen in het business-model nieuwe database-schema's.

Herbruikbaar in plaats van éénmalig.

Gebruik zoveel mogelijk gestandaardiseerde en herbruikbare



Afbeelding 1: Conceptueel informatielogistiek model.



Afbeelding 2: Masterdata management op hoog niveau.

processen. Dit is volkomen logisch, maar in de hectiek en onder de druk van tijdig opleveren, wordt aan dit aspect nogal eens niet de aandacht gegeven die het eigenlijk verdient. Het resultaat is dan dat toch weer 'gauw even' voor dit specifieke geval iets aparts gemaakt wordt. Dit is toch ook vaak een gevolg van het niet voldoende aandacht geven aan het ontwerpen van het initiële proces, hetzij door onvoldoende specifieke kennis, hetzij door wederom tijdsdruk bij het eerste datawarehouse increment.

Ontkoppelen in plaats van integreren.

Op het eerste gezicht lijkt dit niet bevorderend voor veranderingsbestendigheid. Immers het idee van een geïntegreerde omgeving is toch dat juist door deze integratie minder veranderingen nodig zijn. Het is echter beter de verschillende informatie-logistieke functies te ontkoppelen en in aparte lagen onder te brengen. Veranderingen blijven dan – net als in de fysieke logistieke processen – beperkt tot één of enkele specifieke functies; 'het indammen van de olievlek'. Als daarnaast dan het ontwerp van een individuele functie zoveel mogelijk generiek of gestandaardiseerd is, is het des te makkelijker het effect van veranderingen te isoleren en daarmee de doorlooptijd van het doorvoeren van veranderingen te minimaliseren.

Federatief in plaats van centralistisch.

Bij het ontwerpen van een enterprise datawarehouse is het gebruikelijk om uit te gaan van één centrale implementatie, waarin op één plaats in de organisatie alle gegevens (zowel de feiten als de masterdata) bijeengebracht en gemanaged worden, en waar vanuit de diverse datamarts ververst worden. In de

meeste situaties is het beter om een federatief concept toe te passen. In een federatief concept wordt het enterprise datawarehouse opgesplitst in een aantal samenhangende en goed op elkaar afgestemde deelimplementaties. Iedere afzonderlijke business unit (regio, land of werkmaatschappij) met een eigen bestuurlijke verantwoordelijkheid, krijgt een eigen implementatie waarin de business unit zelf de vrijheid en verantwoordelijkheid heeft de masterdata van 'zijn' deel van het business-model in te richten en te onderhouden. Op deze wijze is het makkelijker om draagvlak te krijgen in 'autonome' delen van de organisatie en hoeft ook 'alleen maar' afstemming plaats te vinden over datgene wat gezamenlijk gebruikt wordt.

Erfenis

Een initiatief om een enterprise datawarehouse te ontwikkelen vindt praktisch altijd plaats op het moment waarop er – afhankelijk van de totale omvang van de organisatie – vele tientallen, misschien wel honderden systemen en systeemjes bestaan die management-informatie leveren. Lokale systemen, afdelings-systemen, systemen op de PC van individuele gebruikers (60 procent van alle management-informatie wordt gemaakt met Excel) enzovoort. En ook hier geldt 'dat de geest willig is, maar het vlees zwak'. Iedereen is het erover eens dat zo'n situatie voor niemand goed is, maar als puntje bij paaltje komt zijn toch een heleboel mensen niet bereid hun autonomie op te geven. Centrale financiering van de basisinfrastructuur en een federatief concept zoals hiervoor beschreven is de beste manier om met dit probleem om te gaan. De voordelen van een gezamenlijke en gedeelde infrastructuur moeten groter zijn dan de (gepercipieerde) nadelen. Kortom, het is geen kwestie van opleggen van een centraal (liever gezamenlijk) datawarehouse maar van overtuigen en verkopen van het nut ervan.

Data

En dan is alles geregeld zoals hierboven beschreven. Geld en governance zijn op de juiste wijze belegd, het ontwerp van het datawarehouse en de logistieke processen voldoet aan de regels, de ingezette technologie is *state of the art*, er wordt rekening gehouden met de gevoeligheden van de organisatie met betrekking tot de bestaande systemen en toch ... Na de aanvankelijke successen die waarschijnlijk gevierd kunnen worden, stopt de ontwikkeling. Naarmate integratie van gegevens uit bronsystemen beter geregeld kan worden, kunnen er ook meer en meer geïntegreerde rapportage- en analysefaciliteiten aangeboden worden. En dan blijkt al snel dat de kwaliteit en consistentie van gegevens vaak (veel of zeer veel) te wensen overlaten. Dit gebrek aan kwaliteit kent twee aspecten. Allereerst fouten in de bronsystemen. Heel veel bronsystemen bevatten foutieve gegevens. Dit kan allerlei oorzaken hebben. In oude operationele systemen zijn bepaalde invoervelden in de loop van de tijd voor heel andere gegevens gebruikt gaan worden dan oorspronkelijk bedoeld was. Vaak moeten gebruikers ook gegevens vastleggen, die voor de eigen functie eigenlijk helemaal niet van belang zijn.

De controle op juistheid van dergelijke gegevens laat dan nog wel eens te wensen over. Maar er kunnen ook domweg fouten in de controleprocedures of in berekeningen zitten. Vaak komt dit soort fouten pas aan het licht als er uitgebreidere analysefaciliteiten beschikbaar komen. De omvang en ernst van problemen van fouten in brondata worden in eerste instantie vaak genegeerd of schromelijk onderschat.

En omdat niets menselijks de gebruikers van de rapportage- en analyse-faciliteiten vreemd is, zal – zeker in eerste instantie – het datawarehouse de schuld krijgen, immers dat is – in de ogen van de gebruikers – de leverancier van de gegevens. Het analyseren van de fouten, het opsporen van de uiteindelijke reden ervan, het corrigeren en het in de toekomst verder voorkomen ervan vraagt tijd, geduld en ook relatief hoge (aanvullende) investeringen, ook in de bronsystemen!

In de wat grotere organisaties is standaard nooit standaard

Een tweede datakwaliteitsprobleem is de consistentie van de masterdata binnen en zeker ook tussen de verschillende systemen en onderdelen van de organisatie. Een onderzoek van The Data Warehouse Institute (TDWI) uit oktober 2006 geeft aan dat 83 procent van de respondenten problemen heeft (gehad) door slechte masterdata en dat 54 procent voordelen heeft genoten door goede masterdata. Vergelijkbare percentages, respectievelijk 81 en 54 procent, worden genoemd voor problemen of successen met rapportage en andere BI-functies. Natuurlijk kunnen er net als bij 'gewone' gegevens ook fouten in de masterdata zitten. Maar veel vaker ligt het probleem veel dieper. In grote organisaties is het eigendom van begrippen als 'klant', 'product', 'omzet', 'kalender' en allerlei andere kernbegrippen evenzeer versplinterd

als de organisatorische zeggenschap een de ondersteunende systemen. En hoewel alle individuele systemen en de rapportages voor de (directe) gebruikers van zo'n silo-applicatie waarschijnlijk correct zijn, wordt op het moment dat hoger in de organisatie vragen gesteld worden als 'wat is de winstgevendheid van de verkopen van de 20 best verkochte producten aan de 20 grootste klanten over het laatste kwartaal' de verwarring groot. En ook hiervan krijgt – in eerste instantie – het datawarehouse de schuld. Het oplossen van dit probleem is echter nog vele malen complexer dan het corrigeren van 'gewone' fouten. Masterdata management is in eerste instantie een organisatorische kwestie. Wie is eigendom van welk element van masterdata, wie heeft bij het opstellen van definities welke inbreng, welk bronsysteem is leidend bij de initiële vastlegging en welke ander bronsystemen zijn volgend en hoe zorgen we dat in het datawarehouse de 'golden copy' voorhanden is. Het enterprise datawarehouse is dus misschien wel de plaats waar het probleem zichtbaar wordt, maar het oplossen ervan is een verantwoordelijkheid van de organisatie, waarbij overigens het enterprise datawarehouse weer wel een belangrijke faciliterende rol kan hebben. Een masterdata management-proces op hoog niveau wordt in afbeelding 2 getoond.

Conclusies

Het succes van een enterprise datawarehouse kan afgemeten worden aan het 'verdwijnen' van allerlei eigen systemen en aan de kwaliteit en het gebruik van de informatie die er in zit. Het implementeren zal een lange en moeizame weg zijn met heel veel valkuilen, maar succes is niet onmogelijk. Dit artikel pretendeert zeker niet om een kant klaar recept te zijn, maar wel om een nadere belichting te geven op een aantal ingrediënten en het gebruik ervan.

Ad Stam (ad.a.stam@atosorigin.com) is Principal Consultant bij Atos Origin.

BI-matrix

De BI-matrix geeft u een overzicht van leveranciers op het gebied van Business Intelligence op de Nederlandse markt en werd voor u samengesteld door Paul van der Linden, senior consultant DWH en BI bij Atos Origin.

In DB/M 5 drukten wij een verkleinde vorm van de BI-matrix af.

Vanaf nu bieden wij u de uitgebreide vorm van de BI-matrix aan op onze website www.dbm.nl.

De BI-matrix on-line is een doorzoekbare database, waar u zelf kunt selecteren welke gegevens u van welke BI-leverancier naast elkaar wilt zien. Handig als u een shortlist van bij u passende leveranciers wilt samenstellen.



De BI-matrix is een initiatief van

