

Organisaties zijn vaak niet meer in staat de overvloed aan -interne- informatie te relateren en te interpreteren. Enterprise Warehouse biedt de oplossing voor een information overload.

# information

CENTRAAL BOEKHUIS

## Enterprise warehouse architectuur

Business Intelligence must love it

Door Ing. Emiel van Bockel MMI

*Hedendaags wordt over business intelligence veel geschreven en gesproken. Vaak vanuit een procesmatig perspectief, waarbij gegevens worden 'gepimped' tot informatie om bedrijfsprocessen te ondersteunen en de onderneming te laten excelleren. Daarbij wordt minder vaak melding gemaakt van hetgeen waaraan business intelligence haar bestaansrecht heeft ontleend: het datawarehouse. En juist het datawarehouse moet onder architectuur zijn gebouwd om business intelligence tot grote hoogtes te kunnen brengen.*

Architectuur beschrijft belangrijke keuzes en afwegingen die vooraf zijn gemaakt en leiden tot een gedegen basis voor de lange termijn. De kans dat vervolgens een business intelligence traject faalt wordt daarmee aanzienlijk kleiner. Een belangrijk onderdeel van

deze architectuur is in definitie de wijze waarop het datawarehouse een bijdrage levert aan het realiseren van de doelen die de organisatie zich stelt. Business Intelligence en daarmee het datawarehouse moeten dan ook vanuit strategisch belang zijn geboren.

### “Information overload”

Vandaag de dag kunnen organisaties niet meer zonder informatie. Informatie is in verschillende afdelingen en op verschillende niveaus aanwezig. Individuen en/of afdelingen sturen aan de hand van zelf gefabriceerde lijsten met daarop hun eigen interpretatie van de waarheid. Vaak is bedrijfsbreed de samenhang tussen de onderlinge informatiebronnen zoek. De organisatie is niet meer in staat deze overvloed aan informatie onderling te relateren en te interpreteren. Het wordt daardoor steeds moeilijker de

juiste beslissingen te nemen. Informatie implodeert als het ware in gegevens, hoewel het tegenovergestelde effect moet worden bereikt.

Vaak worden hierdoor nog meer lijsten gegenereerd en de organisatie komt in een neergaande informatie spiraal terecht. De “information overload” is daar. De oplossing ligt bij een centrale omgeving van waaruit eenduidige, consistente maar vooral ook toegankelijke informatie kan worden verstrekt: het “enterprise warehouse”.<sup>1</sup>

<sup>1</sup> Note:

*vaak wordt het begrip datawarehouse gebruikt in de overkoepelende context van operational datastore, extractie-, laad- en transformatieprocessen, datawarehouse, datamarts en meta-data. Om begripsverwarring te voorkomen gebruik ik voor de overkoepelende context de term enterprise warehouse.*

### Datamart versus datawarehouse

Datawarehouse en datamart zijn twee basisbegrippen die worden gebruikt als synoniem voor de centrale informatiebron van een onderneming. Belangrijk is deze ook te definiëren. Het is makkelijk een definitie over te nemen uit de literatuur. Echter, de praktijk leert dat vaak een eigen definitie wordt gehanteerd. In basis wordt de enterprise warehouse-architectuur opgebouwd vanuit de volgende definities:

**Datawarehouse:** een gegevensverzameling voor informatieverstrekking over onderwerpen van de business en is faciliterend naar de gehele organisatie. Een centrale organisatie is hiervan eigenaar.

**Datamart:** een gegevensverzameling over een bepaald onderwerp specifiek voor een afdeling (single line of business) voor beslissingsondersteuning van die afdeling. De afdeling is eigenaar.

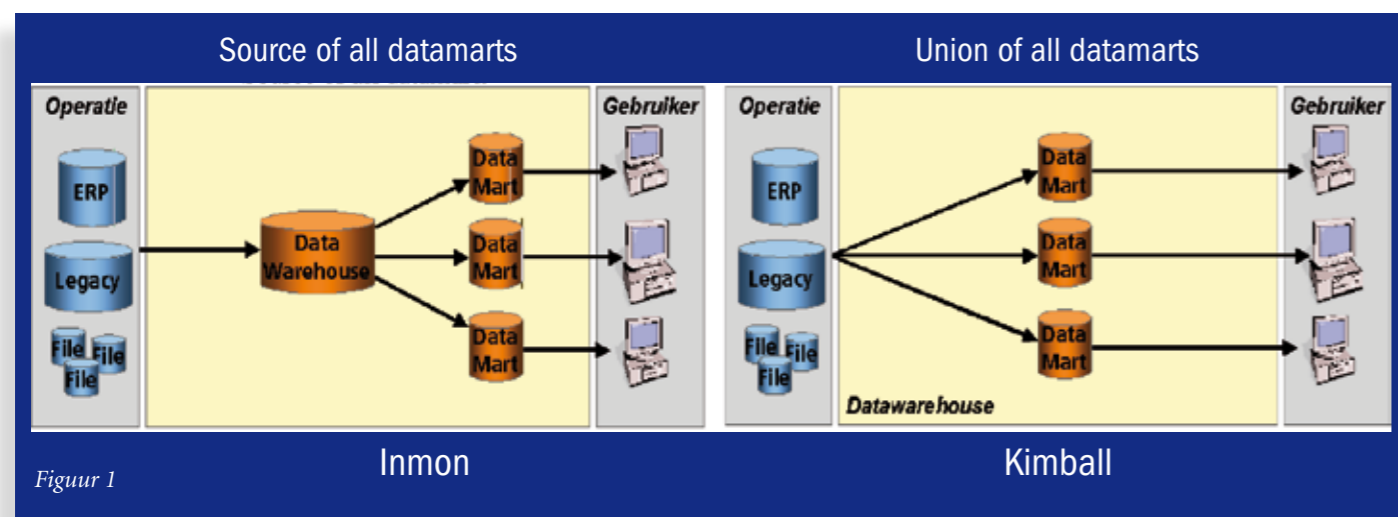
# Centraal Boekhuis

## Vervolg



Deze definities kunnen op twee manieren worden geïmplementeerd. In de literatuur uit zich dit vooral in het verschil tussen de goeroes Kimball en Inmon. Kimball ziet een datawarehouse als 'the union of all datamarts' en volgens Inmon is een datawarehouse 'the source of all datamarts'. Deze opvattingen zijn weergegeven in figuur 1.

vanuit de operationele systemen naar het datawarehouse. In het datawarehouse worden de bedrijfsgegevens eenduidig, uniform en consistent vastgelegd; in optima forma het zuivere corporate datamodel. Vervolgens kunnen datamarts op redelijk eenvoudige wijze worden gewijzigd of uit het datawarehouse worden ontsloten. Aanpassingen in het primaire laadproces,



Figuur 1

Voor Kimball kan worden gekozen als er op eenvoudige en snelle wijze enkele datamarts moeten worden gerealiseerd. Het nadeel van deze keuze is dat overzichtelijkheid en flexibiliteit afnemen bij latere veranderingen. Nieuwe datamarts met nieuwe laadprocessen vanuit de operationele systemen zullen ontstaan en daardoor wordt het steeds complexer om de gegevens tussen de datamarts onderling consistent te houden. Dit laatste is noodzakelijk om te voorkomen dat de samenhang wederom zoekraakt. Naarmate het aantal datamarts toeneemt, neemt het overzicht af en de neergaande informatiespiraal is weer ontstaan.

Een betere keuze is de centrale positie van het datawarehouse. Daarmee is er één centraal laadproces

operationele systemen naar het datawarehouse, of andere datamarts zijn daarbij niet nodig. Na deze afwegingen te hebben gemaakt, is het noodzakelijk

de logische grondslag te definiëren waarop de informatiearchitectuur van zowel het datawarehouse als de datamart berust:

DATAWAREHOUSE	DATAMART
Faciliterend naar de gehele business	Faciliterend naar één single line of business
Gegevensmodel is een reflectie van de business	Gegevensmodel is een reflectie van de vragen die spelen op de afdeling
De data is relationeel gemodelleerd	De data is dimensioneel gemodelleerd
Doorgaans genormaliseerd	Doorgaans gedegenormaliseerd
Detailinformatie	Geaggregeerde informatie
Minimaal geïndexeerd	Maximaal geïndexeerd
Omgeving voor professionals	Omgeving voor gebruikers
Stabiel qua verandering	Flexibel qua verandering

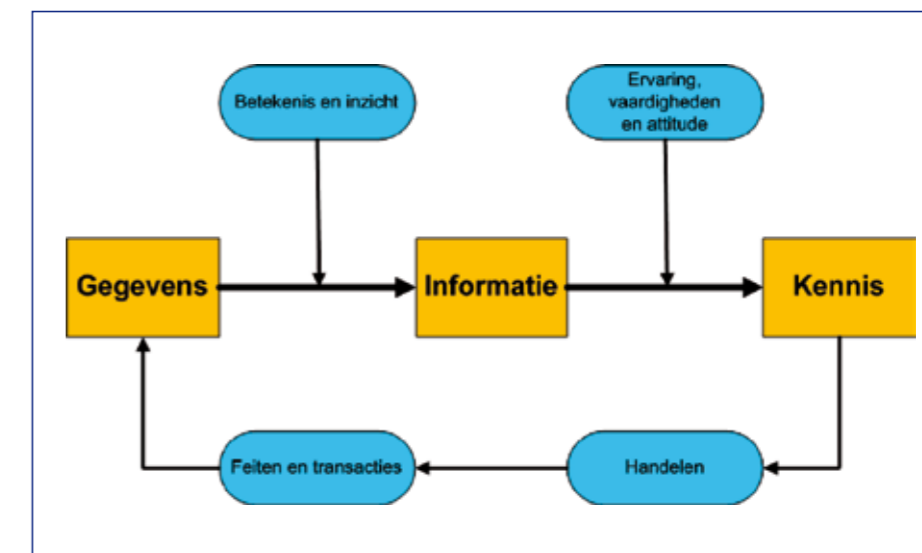
Elke verandering binnen het enterprise warehouse moet in lijn zijn met de uitgangspunten die in de architectuur zijn gedefinieerd. Een afwijking hierop moet goed worden overwogen en eenduidig worden vastgelegd. Hierin zit namelijk op langere termijn precies het verschil tussen de vruchten plukken of de prijs betalen.

Nu conceptueel helder is waaraan de gegevensstroom moet gaan voldoen, is het zaak het warehouse te gaan vullen met gegevens. Daarbij is de vraag welke gegevens?

**Gegevens, Informatie en Kennis**  
Bij business intelligence worden de woorden kennis, informatie en gegevens vaak door elkaar gebruikt. Toch is er een essentieel verschil tussen de drie begrippen. Om dit verschil duidelijk te maken wordt gebruik gemaakt van een stuk theorie uit de wereld van kennismanagement.

Gegevens zijn niets meer dan een set aan discrete en objectieve feiten van bepaalde handelingen die hebben plaatsgevonden. In organisatorische context zijn gegevens de records die zijn voortgekomen uit transacties. Voorbeelden zijn geplaatste orders, distributieopdrachten en vorderingen. Over het algemeen worden deze gegevens vastgelegd in een relationeel databasesysteem. Informatie is afgeleid van het woord informeren, wat betekent inzicht

verschaffen of betekenis gegeven. Door iemand informatie te geven ontstaat een bepaald inzicht. Met dit inzicht is de organisatie in staat handelingen te verrichten. Informatie leidt als het ware tot actie. Als dit niet het geval is, was de informatie voor de persoon die de informatie heeft ontvangen niets meer dan een gegeven. Het is dan ook zeer persoonsgebonden en afhankelijk van de waarde die er aan wordt gehecht of iets een gegeven is of informatie. Aan informatie wordt een mix van ervaring, vaardigheden en attitude toegevoegd waardoor kennis ontstaat. Binnen een organisatie leidt kennis tot handelen waardoor nieuwe feiten en transacties ontstaan. Hierdoor ontstaat als het ware een "Closed-loop cycle of information" (CLCI). Grafisch kan dit als volgt worden weergegeven:



**Je weet niet wat je wil weten**  
Het klinkt wellicht simpel zo'n "Closed loop of information". Echter, het probleem is dat geen enkele gebruiker binnen de organisatie vooraf weet wat hij achteraf wil weten. Afhankelijk van beschikbare informatie en inzichten zullen altijd weer (door opgedane kennis) nieuwe vragen ontstaan. Het is onmogelijk om het hele scala aan vragen binnen een organisatie in kaart te brengen. De focus moet liggen op de belangrijkste vragen. Deze sluiten aan bij de onderwerpen waar het management over praat. Notulen van een managementoverleg zijn goede bronnen ter verificatie.

Het is belangrijk het doel van een enterprise warehouse nooit uit het oog te verliezen, namelijk ondersteuning van de organisatie om doelen te realiseren. Om doelen te realiseren moeten beslissingen worden genomen, gebaseerd op waardevolle informatie. Een gebruiker wil de beschikking hebben over alle gegevens, tot in het laagste detail. Bij elke vorm van informatie moet de analist zich afvragen welke business beslissing (lees handelen in de CLCI) aan de hand van de informatie kan worden genomen. Is hierop door de organisatie geen antwoord te geven, dan is de informatie waarschijnlijk niets meer dan een gegeven. Het is dan de vraag of het gegeven überhaupt wel in het datawarehouse deel van het enterprise warehouse moet worden opgenomen, laat staan in de uiteindelijke datamart.

## Centraal Boekhuis

Vervolg

### Gebruikers

Veel BI-leveranciers schrijven over de mooiste toepassingen van on-line analytical processing (olap). Gebruikers worden gecharmeerd met “Balance score cards” en “Business Activity Monitors” die onder de noemer van “Corporate Performance Management” kunnen worden geïntroduceerd. Echter, veel gebruikers van een enterprise warehouse zijn geen olap'ers of weten niet welke acties moeten worden ondernomen als gevolg van meterstanden op hun dashboard. Zij zijn al tevreden als ze aan de hand van eenvoudige en overzichtelijke rapportages beslissingen kunnen nemen.

Zo heeft Centraal Boekhuis onlangs de publieksprijs van de BI Awards 2006 in de wacht weten te slepen voor haar online applicatie verslaggeving, management en marktinformatie (VMM)<sup>2</sup> ten behoeve van vijfhonderd uitgevers in Nederland. Het publiek gaf als reden: duidelijk wat de uitgever met de informatie kan en overzichtelijk, de informatie is laagdrempelig en toegankelijk voor iedere gebruiker binnen de uitgeversorganisatie. Dit informatieportaal bevat een aantal eenvoudige overzichtelijke maar ook flexibele rapportages dat de primaire bedrijfsprocessen van een uitgever ondersteunt. Tijdens het gehele project was de enterprisewarehouse architectuur het fundament met als belangrijkste pijler het tonen van informatie waarmee ook beslissingen kunnen worden genomen.

#### <sup>2</sup> Note:

VMM is een BI applicatie waarmee de uitgever beter in staat is voorraadniveau te optimaliseren, verkopen te volgen en te analyseren. Ook wordt inzicht gegeven in zaken als distributiekosten, de afdrachten en performance van klanten (boekverkopers), waardoor de uitgever beter in staat is met boekverkopers in onderhandeling te treden en kosten te verminderen. Als laatste marktinformatie waarmee een uitgever in staat wordt gesteld beslissingen te nemen over het starten of stoppen van fondsen en/of titels gezien de bewegingen in de markt.



### De kosten

De meeste kosten van een enterprisewarehouseproject liggen bij het vullen van het datawarehouse. Tijdens deze fase komen complexe onderwerpen aan bod, zoals het corporate datamodel en de inrichting van de extractie-, transformatie- en laadprocessen. Het is verleidelijk om hieraan weinig aandacht te besteden omdat de organisatie wacht op de realisatie van haar datamart. Er bestaat daardoor een grote kans in de eigen valkuil te trappen, namelijk dat in het datawarehouse dan beslissingen worden genomen op basis van de wens van de afdeling waarvoor de datamart wordt ontwikkeld en niet op basis van het belang van de gehele organisatie. Dit druist in tegen de uitgangspunten zoals die zijn gedefinieerd in de informatiearchitectuur. Het resultaat van zulke beslissingen zal niet direct zichtbaar zijn, maar zal zich uiten tijdens het onderhoud. Vroeg of laat zullen de afdelings-specifieke elementen in het datawarehouse moeten worden omgezet naar uniforme elementen. Investeren in het goed vullen van het datawarehouse verdient zich automatisch terug tijdens creatie van de datamarts. De ervaring leert dat hoe beter het datawarehouse is ingericht, hoe makkelijker datamarts eruit kunnen worden ontsloten. Maar houdt ook hier rekening met de CLCI. Hanteer als vuistregel: stop pas een gegeven in het datawarehouse als duidelijk is welke besluiten genomen of handelingen verricht kunnen worden.

### Conclusie

Business intelligence is een middel om de organisatie te verbeteren en de doelen die de organisatie zich stelt te realiseren. De beschikbaarheid van eenduidige, consistente en waardevolle informatie speelt daarbij een cruciale rol. Het is belangrijk om in het bezit te zijn van een architectuur om het enterprise warehouse in lijn te houden met de informatiebehoefte, en daarmee de intelligentie van de business.

Iedere architectuur begint bij de definitie en een beschrijving van een aantal essentiële onderwerpen op conceptueel niveau. Vervolgens worden de onderwerpen met elkaar in relatie gebracht en wordt de samenhang beschreven. Het is daarbij zeer belangrijk om niet te ver in detail te treden, om het overzicht te behouden. Binnen de architectuur wordt de focus gelegd op de onderwerpen die belangrijk zijn voor de organisatie. Door deze onderwerpen te projecteren op de praktijk, ontstaat snel een beeld hoe de architectuur er op hoofdlijnen uit moet komen te zien. Architectuur moet niet alleen theoretisch blijven maar toepasbaar zijn in de praktijk. Door architectuur te benaderen vanuit de praktijk, wordt het tastbaar en overzichtelijk. Daarmee bewijst het de toegevoegde waarde voor de organisatie.

**Ing. Emiel van Bockel MMI**  
IT-Architect Business Intelligence &  
Datawarehousing / Centraal Boekhuis  
E-mail: e.van.bockel@centraal.boekhuis.nl

# ADVERTENTIE