

Naslagwerk volgegoten met Inmon-kennis

Het pakhuis is een complete fabriek geworden

Paul van der Linden

Sinds zijn meesterwerk *Building the Data Warehouse* heeft Bill Inmon een hele reeks van artikelen en boeken doen verschijnen waarin het datawarehouse en aanverwante onderwerpen centraal staan. De boodschap van 'Building' was dat er een bestaansrecht is voor het datawarehouse. Breder bekeken: Inmon toonde aan dat naast de bekende operationele of transactionele omgeving zoiets bestaat als een datawarehouse-omgeving.

Die datawarehouse-omgeving is inmiddels de corporate information factory (CIF) gedoopt. In het gelijknamige boek van de hand van Inmon, Imhoff en Sousa wordt uitgelegd dat de CIF bestaat uit componenten als de operational datastore (ODS), data-marts, een *exploration warehouse* en een dataminingwarehouse, een integratie- en transformatielaag, DSS-applicaties en een datawarehouse. Niet dat al deze componenten tegelijkertijd aanwezig zijn. Zoiets groeit geleidelijk, waarbij de business leidend is.

"Corporate Information Factory", inmiddels toe aan zijn tweede druk, staat boordevol met informatie die betrekking heeft op allerlei aan de CIF gerelateerde aspecten. Voor wie zich bezighoudt met datawarehouses biedt de inhoud niet veel nieuwe inzichten. Toch is het beslist een boek om aan te schaffen. Het lijkt wel of Inmon hier is leeggelopen, waardoor het boek uitstekend als referentiebron kan dienen. Voor al degenen die niet dagelijks met datawarehouses en ODS'en bezig zijn is het boek ook niet te versmaden. Door het te lezen wordt in betrekkelijk korte tijd duidelijk wat de CIF is en waarom het belangrijk is voldoende hiervan af te weten.

Een nieuw onderwerp is de zogenaamde oper-mart. De oper-mart kan het beste worden gezien als een combinatie van een operational datastore en een datamart. Anders gezegd: het is een kleine ODS, multidimensionaal van opzet, ten behoeve van snelle

analyse. De oper-mart is zo nieuw, dat hij verder niet als aparte component in de CIF is opgenomen. Iets wat in de volgende editie vast wel zal gebeuren. En natuurlijk kunnen we een apart boek over de oper-mart binnenkort verwachten.

Op enkele punten is het boek wat minder sterk dan je zou

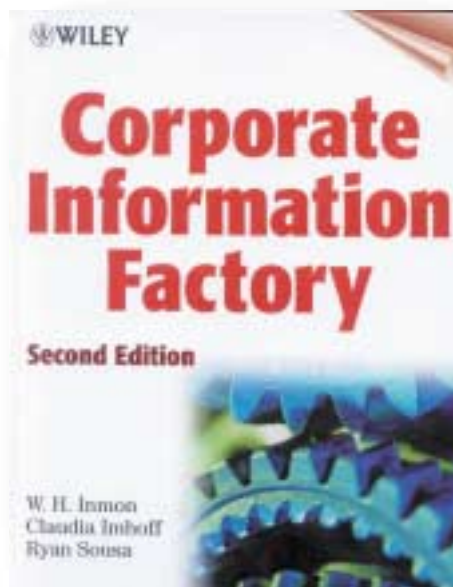
mogen verwachten. Waar Inmon het heeft over databaseselectie komen de *niche strategy* (de rest van de wereld noemt dit de best-of-breed-aanpak) en de *general purpose strategy* aan bod. Inmon stelt dat beide aanpakken in de praktijk op hetzelfde neerkomen. De reden hiervoor is dat bij een general purpose strategy sprake is van configuraties die verschillen per component. Mij lijkt dat overal bijvoorbeeld Oracle inzetten (verschillend geconfigureerd) versus DB2, Red Brick en Tandem niet op hetzelfde neerkomt. Denk alleen maar aan de verschillen in beheer- en opleidingskosten.

Een tweede zwakke plek is Inmons verklaring dat 75 procent van de CIF weinig of geen beveiliging nodig heeft. Een uitspraak die ik niet graag zou overnemen. En ook waar

het eigendom van gegevens aan de orde komt, ontbreekt een grondige behandeling van het onderwerp. Slechts de oppervlakte wordt geraakt. Desalnietemin is het boek een absolute aanrader, met name door de breedte aan onderwerpen die worden behandeld en de manier waarop de lezer inzicht krijgt in de opbouw en werking van de CIF.

W.H. Inmon, Claudia Imhoff en Ryan Sousa,
Corporate Information Factory. John Wiley & Sons, 2001,
ISBN 0-471-39961-2. 400 pag., prijs ca. € 56,-.

Folio wordt vervolgd op pagina 46



Verplichte kost voor het e-bedrijf

"Data Warehousing for E-Business" bouwt voort op hetgeen Inmon en de zijnen eerder over de corporate information factory schreven.

Bijgestaan door Terdeman, Norris-Montanari en Meers snijdt warehouse-goeroe Inmon hier een bijzonder actueel thema aan.

Beschrijft hij in 'CIF' een generieke architectuuropzet, in dit boek wordt behandeld hoe de CIF eruitziet voor een e-business-opzet. De uitdaging bij het inrichten van een e-business-organisatie is dat die potentieel veel gegevens moet kunnen verstouwen, dat processen herhaalbaar moeten zijn, dat snelheid centraal staat en dat de beschikbaarheid van goede metadata essentieel is. Bouw je een e-business-architectuur die onvoldoende schaalbaar en flexibel is, dan zijn de kosten die je achteraf moet betalen formidabel.

Het is zaak zo optimaal mogelijk om te gaan met clickstream-data en personalisatie-engines, om zo de verschillende soorten consumenten (onderzoekers, kijkers, kopers en community-leden) zo goed mogelijk te bedienen. Hierbij is een

belangrijke rol weggelegd voor de metadatatransferten (html, XML en XHTML), maar ook de samensmelting van software voor ETL en messaging wordt gesignaleerd als belangrijke ontwikkeling.

Knelpunten kunnen zich op verschillende punten binnen de e-business-architectuur voordoen.

Men zou dan ook meerdere aspecten continu moeten monitoren. Klinkt logisch, maar gebeurt in de praktijk nog maar zelden. Door het verstrekken van dergelijke tips behoort dit boek dan ook tot de verplichte kost voor e-business-bedrijven.

W.H. Inmon, R.H. Terdeman, Joyce Norris-Montanari en Dan Meers, Data Warehousing for E-Business. John Wiley & Sons, 2001, ISBN 0-471-41579-0. 272 pag., prijs ca. € 50,-.

Drs. P.F.H. van der Linden (PvanderLinden@Synergetics.nl) is principal consultant bij Synergetics, the Management Information Company.

