

Didactische helderheid en ambitieuze volledigheid

Verdienstelijke boeken over BI en datawarehousing

Paul van der Linden

Het boek 'Datawarehousing. Een inleiding' van Karien Verhagen vormt een goede en vlot leesbare introductie in het vakgebied. Het is een heel didactisch geschreven werk, waarin een complexe en omvangrijke materie toch helder wordt gebracht. In tien hoofdstukken en 194 pagina's wordt datawarehousing geïntroduceerd en uitgelegd voor personen die weinig tot geen kennis en ervaring hierin hebben. Denk dan aan studenten die in dit veld werkzaam willen zijn of managers die zijdelings of af en toe te maken hebben met datawarehousing. Voor beide categorieën geldt dat ze in dit boek een goed opstapje zullen vinden.

Vier kanten

Om datawarehousing uit te leggen verdeelt Verhagen het werkveld in een achterkant, een binnenkant, voorkant en zijkant. De achterkant gaat dan over de aanleverende systemen, de binnenkant is het data warehouse zelf en de voorkant heeft betrekking op de presentatie van informatie (gebruikerstoepassingen). De zijkant tenslotte betreft het beheer. Deze opzet is helder en sluit ook goed aan bij de normale manier waarop dit soort plaatjes worden getekend. Wat wel bevreemd is dat het ETL-proces (extractie, transformatie en laden van data) wordt gezien als onderdeel van de zijkant en niet van de achterkant. Wat het boek met name wil is het logische concept datawarehousing uitleggen. Echt technische aspecten en te maken keuzes komen dan ook minder aan bod. Dit is gezien het inleidende karakter van het boek een te verdedigen keuze. Op een aantal plaatsen wordt net iets gemist. Zo wordt het multidimensionele model wel behandeld, maar komt multidimensionele opslag niet aan de orde. Bij onderdeel 5.14 worden speciale manieren van indexering en speciale hardware genoemd als manieren om verbeteringen in het gedrag van de database te bewerkstelligen. Hier zou ik ook speciale databases/software verwachten. Denk dan bijvoorbeeld aan kolomgeoriënteerde databases of aan query-versnellers.

Jammer is dat nergens wordt gemeld dat het data warehouse een toepassingsafhankelijke gegevensverzameling is en dat een datamart een toepassingsafhankelijk karakter heeft. Stermodelleren en Kimball worden wel genoemd maar de busstructuur komt niet aan bod. Zoiets zou wel besproken moeten worden – ook in een introductie tot datawarehousing. Bij de behandeling van de datawarehouse-theorie worden Inmon en

Kimball genoemd, maar bij de uitleg van het relationele model is ervoor gekozen om Date en Codd niet te noemen. Nu verdienen de prestaties van Inmon en Kimball best wel respect, maar vergeleken met Date & Codd zijn het dwergen.

In hoofdstuk 5 gaat het om de keuze van een oplossing (relationeel versus multidimensioneel). Hier wordt het niet noemen van de busstructuur en data staging area pijnlijk. In praktijk wordt vaak gekozen voor een relationele opslagstructuur in de staging area met een conformed dimensions opzet (busstructuur) als datawarehouse. Deze mogelijke oplossing wordt hier dan ook niet besproken.

CD Rom

Wat meestal een teleurstellend toetje is, is bij Verhagen een aangename afsluiter. In de CD Rom die bij het boek hoort wordt een verdere verdieping aangeboden. Hier maken verschillende BI-leveranciers evenals het onvermijdelijke Gartner hun opwachting en worden onderwerpen als CWM (Common Warehouse Metamodel) en FCO-IM aangestipt. Dit alles weliswaar in de vorm van een verzameling van brochures en artikelen, maar de extra informatie zal beginners in het vakgebied zeker niet onwelkom zijn.

Ondanks de eerder gemaakte kanttekeningen wordt 'Datawarehousing. Een inleiding' mede daardoor een solide en prettige inleiding in het werkveld. Een aanrader voor iedereen die zich hierin wil of moet verdiepen.

*Karien Verhagen: Datawarehousing. Een inleiding.
Pearson Education Benelux, Amsterdam (2005).
ISBN 90-430-0989-X.*

Indrukwekkend beeld

Waar Verhagen kiest voor een chirurgische behandeling van datawarehousing heeft het boek van Daan van Beek meer weg van een ouderwets bommentapijt. In eveneens tien hoofdstukken (maar 295 pagina's) wordt een zeer indrukwekkend beeld van de intelligente onderneming neergezet. Om het beeld van een intelligente onderneming vorm te geven gebruikt Van Beek het beeld van de levende organisatie. De intelligente onderneming verwerkt relevante signalen door de betreffende data om te zetten in informatie en kennis en daar vervolgens op te acteren. Business Intelligence (BI) vormt hierbij het instrument om de uitdaging van



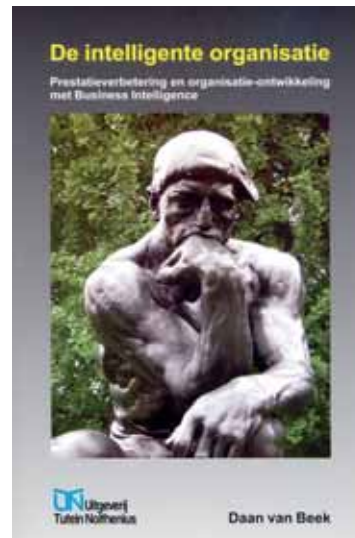
steeds meer data en steeds minder bedenktijd succesvol aan te gaan.

Qua ambitie gaat 'De Intelligente Onderneming' multidimensionaal over het boek van Karien Verhagen heen. Ten eerste is de scope veel ambitieuzer: datawarehousing is een klein onderdeel van hetgeen Van Beek over de intelligente organisatie te vertellen heeft. Maar ook de doelgroep is veel ruimer. Om op te sommen voor wie het boek bedoeld is heeft Van Beek dan ook een halve pagina nodig.

Aanpak

Van Beek is wat stilliger in zijn aanpak: geen uitgebreide behandeling van relationeel versus stermodel. In plaats daarvan wordt het sterschema voor datamart en centraal datawarehouse aanbevolen (busstructuur). Hierover kun je discussiëren, maar belangrijker is dat hier een duidelijke stelling en oplossing wordt neergelegd en niet een aantal mogelijke architecturen wordt afgewogen. Didactiek zit hier duidelijk niet in de driver's seat. Van Beek probeert alle aspecten van de intelligente organisatie aan de orde te stellen. En omdat dat er behoorlijk veel zijn is zijn tempo adembenemend en zijn behandeling kort en bondig. Onderwerpen als BI-ambitieniveaus, stadia van de intelligente organisatie, kritieke succesfactoren, OLAP, metadata repository's, ETL, dashboards, INK, Six Sigma, CRM, forecasting, visualisatie, proceskanteling etcetera komen allemaal langs waarbij niet wordt vergeten om de autoriteiten op deze gebieden even op te voeren. Zo worden rondom het onderwerp organisatie van werk Weber, Mintzberg, Zuurmond en Bekkers aangehaald. Handig voor wie zich daar verder in wil verdiepen (hetzelfde geldt voor de bijgevoegde literatuurlijst).

Niet duidelijk is waarom er in Van Beeks uitputtende boek geen aandacht is voor standaarden als Common Warehouse Metamodel (CWM) of XML/A. Ook Codd en Date worden vergeefs gezocht op de literatuurlijst – ze horen er absoluut op thuis. Wellicht dat in deze gevallen de snelheid gewoon te hoog is geweest. Maar



ondanks deze 'schoonheidsfoutjes' is 'De intelligente organisatie' welbested aan mensen die met BI hebben te maken en eens een breder perspectief op de materie willen krijgen.

Daan Van Beek: De intelligente organisatie. Prestatieverbetering en organisatie-ontwikkeling met Business Intelligence. Tutein Nolthenius (2004). ISBN: 90-72194-72-1.

Conclusie

Zowel Verhagens 'Datawarehousing. Een inleiding' als Van Beeks 'De intelligente organisatie' zijn verdienstelijke bijdragen aan de Nederlandstalige literatuur over BI en datawarehousing. Waar Verhagen uitblinkt in een heldere, duidelijke en didactische behandeling van het brede thema datawarehousing heeft Van Beek geprobeerd om met name volledig te zijn en de aansluiting te zoeken bij relevante literatuur. Het gevolg van deze keuzes is dat het boek van Verhagen zich het beste leent voor mensen die hun eerste stappen in het vakgebied willen zetten. In dat geval is 'Datawarehousing. Een inleiding' zeker een aanrader. Daan van Beeks boek is niet voor de beginners. Het werkt ook eerder als een naslagwerk en springplank tot nog meer (verwijzingen naar andere auteurs en werken zijn legio aanwezig) en past daarmee beter bij mensen die al meerdere jaren kennis en ervaring in het vakgebied hebben opgedaan en zich met name in de breedte verder willen verdiepen. De scope van hetgeen in 'De intelligente organisatie' wordt behandeld is indrukwekkend. Hetzelfde geldt voor de informatiedichtheid. Misschien dat hier het advies geldt: met kleine slokjes. Beide boeken zijn het lezen zeker waard. Slechts door de didactische helderheid van een Karien Verhagen te koppelen aan de ambitieuze volledigheid van een Daan Van Beek wordt een nog hoger niveau bereikt. In dat geval zou sprake zijn van een standaardwerk.

Paul van der Linden (Paul.PFH.vanderLinden@AtosOrigin.com) is senior consultant Data Warehousing/BI bij Atos Origin en geeft leiding aan Data Warehousing Cost & Lifecycle Management (CLM).