

Jill Dyché over trends in datamanagement en data-integratie

Data as a Service

Hans Lamboo

De Amerikaanse Jill Dyché is partner bij de Baseline Consulting Group, gespecialiseerd in datawarehousing, Business Intelligence, data-integratie en datamanagement. Ze was op uitnodiging van SAS enkele dagen in Nederland. Zij signaleert een aantal trends, waaronder Data as a Service.

Datamanagement werd als laatste aan de portfolio van de Baseline Group toegevoegd. "Gezien de propvolle pijplijn een veelbelovend werkgebied. We hebben daarmee direct de eerste trend te pakken", zegt Jill Dyché. "Bedrijven beginnen zich te realiseren dat data echt een *asset* vormen. En dat ze behoefte hebben aan processen en speciale vaardigheden rond die data. Wat we steeds meer zullen gaan zien is *loosely coupling* van de data aan de applicaties. Master Data Management (MDM) en Customer Data Integration (CDI) zijn dus niet alleen maar datawarehousing voor analytics, niet alleen *packaged* applicaties voor een specifiek gebruik van de data, maar zijn nodig om tot Data as a Service te komen. Dat is de trend."

Information Excellence Center

De reden dat Data as a Service nu een trend is en vijf jaar geleden niet, komt door de ontwikkeling van Service Oriented Architecture. "Om bijvoorbeeld customer data te integreren kun je een webservice verschillende locaties laten aanroepen en naar een engine, een CDI-hub, laten gaan om de benodigde data op te halen, ongeacht de bron. De hub bevat geautoriseerde data.

MDM is een soort operationele integratie van reference data

Zo krijg je dus een *échte Single version of the Truth*", stelt Jill Dyché. "De enige overeenkomst met een datawarehouse is dat je nog steeds data moet dupliceren. Maar door MDM wordt de CDI hub het de facto geautoriseerde record systeem voor alle applicaties die die ene kopie van de data nodig hebben." Om het zover te krijgen, zijn speciale maatregelen en vaardigheden nodig, gecentraliseerd in wat zij een Information

Excellence Center noemt. "De mensen in het IEC begrijpen niet alleen de context voor het gebruik van de data, maar ook de data zélf, het datamodel, de metadata, en de relatie met de business", licht ze toe. "Een IEC stroomlijnt de voorziening van verschillende data voor verschillende business doeleinden op enterprise niveau". Zij ziet MDM als een soort operationele integratie van reference data; CDI is MDM voor customer data. Ze constateert dat veel bedrijven roepen met MDM bezig te zijn, maar eigenlijk CDI bedrijven. Ze komen in eerste instantie niet verder dan customer data en kijken misschien volgend jaar eens naar product-data of andere data.

Kwaliteit

"Lang geleden werd begonnen met datawarehousing om te komen tot de Single Version of the Truth. Maar voor de meeste gebruikers is een datawarehouse een *cul-de-sac*, een doodlopende weg, waar de data naartoe gaan om te sterven. Het is de laatste halte van de Data Supply Chain", zegt Dyché. "Maar het is allemaal veel dynamischer. Ik zie MDM als middelpunt, als een soort *hub-and-spoke* waar de data uit verschillende richtingen komen en gaan. Maar er is maar één verblijfplaats waar de data topkwaliteit hebben, cleansed zijn en akkoord volgens de business rules; en die dus de centrale autoriteit op datagebied is. Nogmaals: SOA maakt het mogelijk de data te ontkoppelen. Roep het aan als het nodig is, en dan niet alle data, maar enkel en alleen die nodig zijn. Zo komen we eindelijk af van alle point-to-point verbindingen die we de afgelopen jaren hebben gebouwd."

Het hele hub-concept vraagt om een uiterst strikte aanpak van datakwaliteit. Dyché: "Daar hebben we alles voor ter beschikking, zoals matching-algoritmes om de waarschijnlijkheid van een match te kunnen vaststellen. Er komen klantdata de hub in, daar wordt de individuele klant herkend, en vervolgens wordt vast-



Jill Dyché: "Ik zie MDM als middelpunt, als een soort hub-and-spoke."

gesteld of er al data van dezelfde klant in de hub aanwezig zijn, die twee records worden samengebracht en zo wordt één geautoriseerd record verkregen".

Data Governance

Om MDM op de juiste manier toe te passen is Data Governance onontbeerlijk, ook een trend volgens Dyché. "Governance heeft minder met datamanagement van doen, het is meer het bepalen van een datapolicy. Daar moet ook de business bij betrokken worden, om antwoord te krijgen op vragen als 'wat beschouwen we als een klant' en 'wat zijn de criteria om een individuele klant te identificeren'. Als daar bijvoorbeeld uitkomt 'een klant heeft altijd een accountnummer en een afleveradres' komen klantdata waar minstens één van beide aan mankeert, nooit in de hub. Met MDM dwingen we data-integriteit af."

Een belangrijk verschil met het datawarehouse, vindt Dyché: "De data in het datawarehouse zijn net zo goed als de ETL is. Of beter: net zo goed als de ETL-programmeur, wiens kennis van de business rules vaak beperkt is. MDM is op vele fronten een doorbraak. De loosely coupling geeft de ontwikkelaars meer tijd te doen waar ze goed in zijn, zonder dat ze data-experts hoeven zijn. Het maakt dat beslissers zich geen zorgen

hoeven maken over de kwaliteit van de data die ze gebruiken." MDM heeft erg veel ingebouwde logica. Het gaat niet alleen om het opslaan van een record, maar er wordt door middel van SOA gekeken waar het record zich bevindt. Er is functionaliteit die het record evalueert, standaardiseert en vergelijkt met andere records, en teruglinkt naar systemen die om het record gevraagd hebben. "Er zit dus veel meer *embedded intelligence* in dan in een datawarehouse dat alleen maar passief wacht op de dingen die komen. Toch claimen veel datawarehouse-beheerders dat ze aan MDM doen. Ten onrechte, want het feit dat er masterdata worden gebruikt in het datawarehouse heeft niets met MDM te maken. Bovendien is een datawarehouse gebouwd voor één enkel doel, is niet in staat met services om te gaan en beschikt niet over de ingebouwde logica die ik net noemde".

Jill Dyché ziet nog geen rol weggelegd voor MDM als het gaat om ongestructureerde data. "Masterdata worden aangeroepen door een scala aan applicaties en zijn dus gecentraliseerde, functionele en operationele data. Dat past nog niet in de ongestructureerde wereld. Als ik in een billing-systeem een nieuwe klant wil invoeren, is het belangrijk te weten of de klant al bestaat. Daarvoor heb je de masterdata. In dat proces spelen de notities die een medewerkster van het callcenter heeft gemaakt geen enkele rol. Belangrijk in termen van CRM, maar niet voor MDM. Je zult ongestructureerde data dan ook niet tegenkomen in een CDI hub – maar mogelijk wel in een datawarehouse, waar het een component kan zijn van de Business Intelligence."

Maturity

Jill Dyché is jurylid van de TDWI Best Practice Awards. Vijf jaar geleden werd daarbij vooral gekeken naar de grootte en performance van het datawarehouse, en of het real-time was. Daarna werd het anders. De bedrijven die de prijs kregen toegekend, scoorden meestal hoog op 'dit waren de eisen en wensen van de business, zo hebben we die vertaald in data-vereisten en dit is het probleem dat we hebben opgelost.'

Er zit in MDM veel meer embedded intelligence dan in een datawarehouse

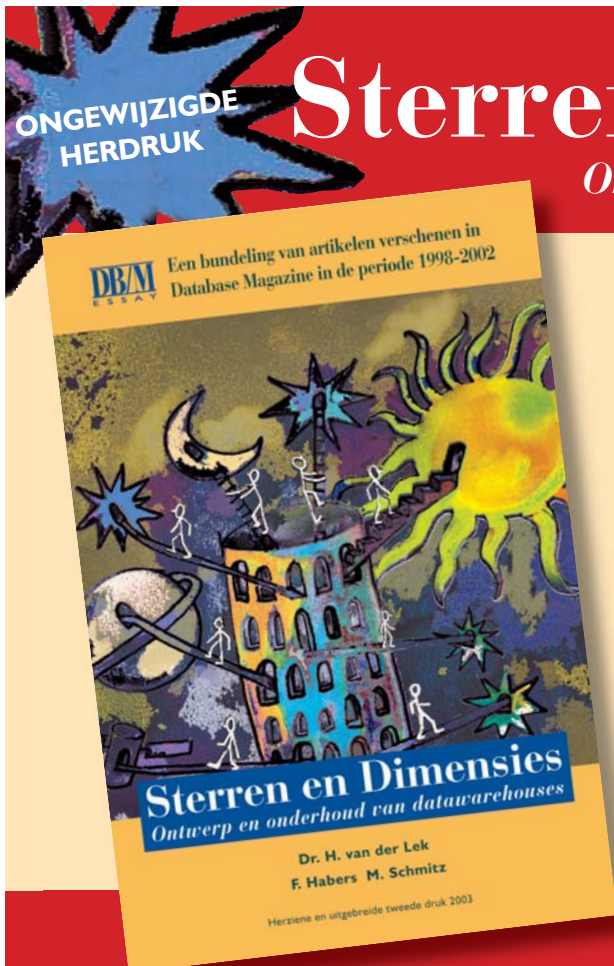
Tegenwoordig gaat dat nog verder: 'dit probleem hebben we opgelost, we staan dichterbij onze klanten dan ooit, we zagen een stijging in klanttevredenheid die leidde tot 53 procent omzetstijging over drie jaar, goed voor 30 miljoen dollar.' Voor het eerst zegt Jill Dyché in gesprekken met klanten: "U zult nooit iets terugzien van uw investeringen in uw datawarehouse. Bij de start beschikte u al niet eens over de goede business case. U heeft veel te veel geïnvesteerd in technologie. Een voorbeeld. Een van mijn klanten heeft een datawarehouse van 10 TB met een stuk of drie gebruikers. Misschien doen die gebruikers wel

sensationele dingen, misschien besparen ze wel veel geld, maar er is 20 miljoen gespendeerd aan het datawarehouse. Het aantal gebruikers is op zich nooit een criterium – ze kunnen wel de top-20 van klanten bedienen – maar dat is hier niet het geval. Het bedrijf heeft een enorme servicegerichtheid naar de klanten, maar gek genoeg mogen belanghebbenden niet hun eigen rapportages draaien. Ze krijgen ze op verzoek per mail. Niet efficiënt, niet schaalbaar en draagt helemaal niets bij aan het verkrijgen van een cultuur gericht op het strategisch gebruiken van informatie door de business. Het betreffende bedrijf krijgt dat binnenkort goed voor de kiezen. En daar praten we momenteel met ze over”. Zij ziet duidelijk correlatie tussen de mate waarin rapportages beschikbaar zijn en de behendigheid van de business om nieuwe acties rond klanten en producten uit te voeren. De waarschijnlijkheid dat een product manager bij een bank aan de hand van zijn verkopen besluit een product aan zijn pakket toe te voegen om het zo beter en sneller te kunnen vermarkten, acht zij recht evenredig met de mate waarin hij toegang tot die informatie heeft. “Als ik bij een klant binnenkom leg ik het bedrijf langs een speciale BI-meetlat en geef hem vervolgens een rapportcijfer. Zie ik binnen zo'n bedrijf iemand die de telefoon pakt als hij informatie nodig heeft, dan bewijst dat dat datawarehousing daar nog in de kinderschoenen staat”.

Hans Lamboo is hoofdredacteur van Database Magazine.

Vier datatypen

Jill Dyché definieert vier verschillende typen data: “Dat zijn masterdata, metadata, reference data en transactiedata. Laten we daar eens naar kijken vanuit business perspectief, bijvoorbeeld een bank. Ik ga naar een geldautomaat om geld op te nemen. Dat is een transactie. Voor de bank is die transactie een event. Reference data verwijzen naar dat event. Dus: ik ben mevrouw die-en-die, ik woon daar, ik heb drie rekeningen, ik kom altijd bij dat ene filiaal. Een transactie genereert point-in-time data; er is maar één zo'n event. Ik kan meerdere transacties doen, maar elke transactie is één afzonderlijk event. Reference data kunnen op een aantal plaatsen voorkomen. Mijn gegevens zijn hetzelfde, of ik nu geld uit de automaat haal, online bankier of aan de balie van een filiaal geholpen word. Reference data zijn alomtegenwoordig in de hele organisatie. Als ik zaken doe kunnen de reference data worden gedupliceerd. Als transactie-data worden gerepliceerd, gebeurt dat met een doel. Het is meestal veel specifiek, beperkter en makkelijker te begrijpen. In zijn meest eenvoudige vorm beschrijven metadata de informatie die uit de transactie en de reference data komt. Metadata geven aan hoe het geld op mijn rekening gekomen is, bijvoorbeeld als automatische salarisoverboeking. Dus kunnen de metadata de regels voor automatische salarisoverboeken tonen. Metadata geven een beschrijving over de data. Masterdata zijn eigenlijk hetzelfde als reference data., maar dan cleaned, geharmoniseerd, conform de business rules, geautoriseerd en centraal beschikbaar gesteld”.



Sterren en Dimensies

Ontwerp en onderhoud van datawarehouses

Sterren en Dimensies is vanwege grote belangstelling in een ongewijzigde derde druk verschenen. Het boek uit de welbekende DB/M Essay reeks bevat een bundeling van artikelen uit DB/M over het ontwerpen en onderhouden van datawarehouses. Deze artikelen zijn gepubliceerd in de periode 1998 – 2002. De experts Harm van der Lek, Frank Habers en Michael Schmitz geven principes voor het gebruik van sterschema's en laten zien hoe de 'sterren' uitblinken in eenvoud.

Wilt u de inherente kracht van het dimensionale denken volledig benutten? Dan kunt u niet zonder dit boek!
Ga snel naar www.array.nl en bestel Sterren en Dimensies!

Array PUBLICATIONS