

VAN DER LANS

Misverstanden over het DDP



Vorig jaar heb ik het *Data Delivery Platform* (DDP) geïntroduceerd; zie de eerste artikelen op www.BeyeNetwork.com. Het DDP is een flexibele architectuur voor het ontwikkelen van Business Intelligence systemen. Karakteristiek aan het DDP is dat de datawarehouse architectuur gedeeltelijk gevirtualiseerd wordt. Zogenaamde data consumers, ofwel de rapportage en analytische producten, zijn hierbij losgekoppeld van de data stores, zoals het datawarehouse, de operational data store en de datamarts. De ont koppeling wordt uitgevoerd door een softwarelaag die tussen de data consumers en de data stores geplaatst is. Deze laag geeft de consumers het idee dat er één geïntegreerde database benaderd wordt, terwijl er technisch waarschijnlijk een heterogene verzameling data stores geraadpleegd wordt.

Met de huidige generatie federation servers, zoals Composite Information Server, Informatica Data Services, Dataflux Data Federator en het Denodo Platform, is het vandaag de dag mogelijk een dergelijke softwarelaag en dus architectuur te ontwikkelen.

Tijdens lezingen en gesprekken merk ik wel eens dat er misverstanden bestaan over het DDP. In deze column wil ik enkele van deze misverstanden proberen op te lossen.

Het meest voorkomende misverstand is dat er gedacht wordt dat het DDP het datawarehouse vervangt, ofwel een organisatie kiest voor een DDP óf een datawarehouse. Dit is absoluut niet het geval. Er kunnen verscheidene redenen bestaan waarom een datawarehouse gewoonweg onontbeerlijk is. Veronderstel dat in bepaalde productiedatabases geen historie bijgehouden wordt, maar dat deze wel nodig is voor trendanalyse. Ergens moeten die historische gegevens dan wel vastgehouden worden en een datawarehouse kan dan een logische plek zijn. Het is ook mogelijk dat het direct uitvoeren van query's op de productiedatabases problemen geeft, bijvoorbeeld dat ze tot teveel verstoring leiden. Een datawarehouse kan dan wederom uitkomst bieden. En zo zijn er meer redenen. Wel schermt het DDP af welke gegevens precies benaderd worden, maar een van de data stores kan wel degelijk een datawarehouse zijn.

Een ander hardnekkig misverstand is dat het DDP ook wel gezien wordt als een virtueel datawarehouse. Het DDP zou slechts toegang geven tot de productiedatabases en er zou

geen enkel datawarehouse of datamart opgebouwd worden. Elke vraag die vanuit de consumers afgevuurd wordt, wordt op de productiedatabases verwerkt. Eigenlijk zou het mooi zijn als dit zou kunnen, maar dit zal in veel situaties niet realistisch zijn. Men zal er zelden aan ontkomen om een datawarehouse en misschien ook een staging area te bouwen. Wel is het de bedoeling om het aantal data stores te minimaliseren en daarmee de architectuur te simplificeren.

Omdat het DDP on-demand gegevens opvraagt, moet er veel on-demand geïntegreerd en getransformeerd worden. Uiteraard kost dit tijd en resources. Sommigen hebben echter het gevoel dat dit altijd teveel kost. Snel wordt de conclusie getrokken dat het DDP alleen ingezet kan worden bij een kleine omgeving, ofwel bij een omgeving met niet al teveel gegevens en gebruikers. Of dit zo is, ligt puur aan de kwaliteit van de software en hardware die ingezet wordt. Door de vele optimalisatiemogelijkheden die de hierboven genoemde federation servers tegenwoordig bieden, zijn ze in staat om grote omgevingen te ondersteunen. Wat ook zeker helpt, zijn de krachtiger servers waarin we enorme geheugens kunnen opbouwen. Misschien is het DDP nog niet klaar voor de allergrootste omgevingen waar het datawarehouse bestaat uit vele Petabytes aan gegevens en waar dagelijks Terabytes aan nieuwe gegevens geladen worden, maar hoeveel organisaties betreft dat nu eigenlijk?

Een vierde misverstand is dat gedacht wordt dat in een DDP gegevens niet on-demand opgeschoond kunnen worden. Dit is niet correct. Ook dit ligt aan de software die ingezet wordt. Bijvoorbeeld, de federation server van Informatica genaamd Data Services kan wel degelijk live de gegevens schoonmaken.

Hopelijk heldert deze column voor sommigen een aantal zaken op en wordt het duidelijker wat het Data Delivery Platform is en wat niet. Het DDP is een realistische architectuur voor het opzetten van Business Intelligence systemen, één die past bij de huidige wensen van gebruikers en die optimaal gebruik maakt van de hardware- en softwaremogelijkheden van vandaag.

Rick van der Lans is zelfstandig IT-consultant.