

Case beschrijft twee datamodelleringopties

Data Vault; to model or not to model?

Fons Pieters

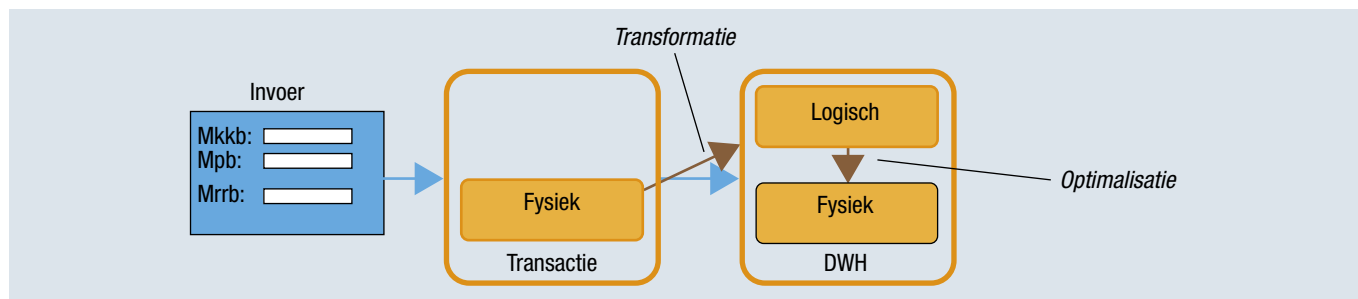
To model or not to model? lijkt een vreemde titel aangezien dit artikel gaat over Data Vault modellering. Maar dit artikel zal aantonen dat deze titel lang niet zo vreemd is als deze lijkt. Sterker nog, het is een vraag die beantwoord dient te worden door de Data Vault specialist voordat wordt begonnen aan een Data Vault gerelateerd project. Een Data Vault model kan namelijk niet alleen worden gemodelleerd op basis van de business-eisen, maar ook getransformeerd (lees afgeleid) vanuit een bronsysteem.

Dit artikel zal duidelijk maken waarom gekozen zal moeten worden tussen het 'modellieren van' een Data Vault model en het 'transformeren naar' een Data Vault model. Laten we beginnen met een praktisch voorbeeld om het geheel te illustreren. Gegeven is een bedrijf in de zware industrie, een eigenaar van meerdere industriële complexen. Asset Management is voor een organisatie als deze zeer belangrijk, het zal toch moeten weten over welke productiecapaciteiten het beschikt. De datahiërarchie die deze organisatie voor een industrieel complex gebruikt is de "Plant" welke opgedeeld kan worden in "Units" waarbinnen zich "Tags" bevinden. De Tag is de functionele locatie waar "Equipment" geïnstalleerd kan worden; denk aan bijvoorbeeld pompen en opslagtanks. Zowel een Tag als Equipment kan beschikken over documentatie; denk aan CAD tekeningen, installatiehandleidingen of veiligheidsvoorschriften. Stel, een niet nader gespecificeerd ERP-systeem wordt geconfigureerd om de gevraagde Asset Management uit te voeren. Het bedrijf bestaat uit meerdere Plants en natuurlijk heeft elke Plant een eigen implementatie van dit ERP-systeem. Documentatie wordt vervolgens ondergebracht in een in eigen beheer ontwikkelde Document Management applicatie terwijl voor de workflow van het beheerproces weer een andere applicatie wordt aangekocht. Vervolgens veranderen de beheerprocessen en begint de busi-

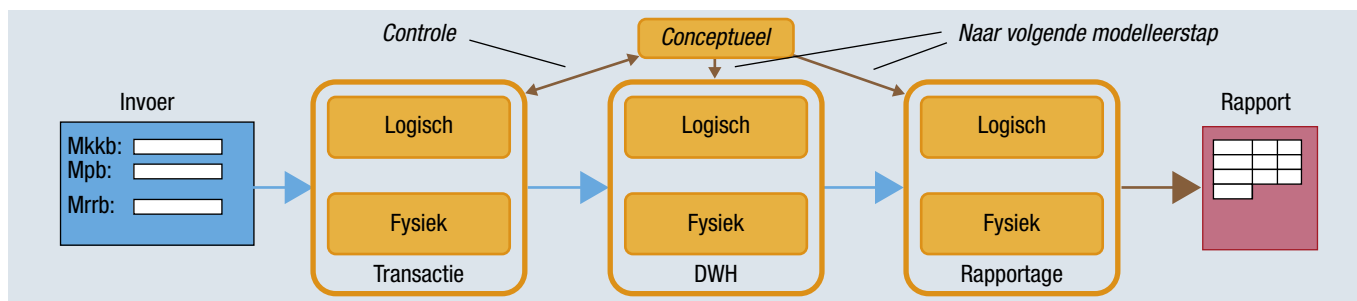
ness de invoervelden te misbruiken om daarmee mogelijk te maken dat ..., u snapt ondertussen waar dit heen gaat.

Transformeren naar een Data Vault

De vraag die nu beantwoord dient te worden voordat kan worden begonnen met het ontwerpen van een Data Vault DWH oplossing (de business case wordt gemakshalve even gezien als een gegeven, inclusief de motivatie voor DV) is "to model or not to model"? In het geval gekozen wordt om niet te modelleren wordt gekozen voor de meest populaire optie om te komen tot een Data Vault. Om deze optie duidelijk te kunnen onderscheiden van de optie "wel modelleren" zal deze transformatie worden genoemd. De transformatie stelt de Data Vault modelleur in staat om een datamodel van een bronsysteem direct te vertalen naar een Data Vault model. Misschien is nu duidelijk waarom dit geen modelleren is maar transformeren, er wordt namelijk geen brug geslagen van de werkelijke businesswereld naar de abstractie van een datamodel. In plaats daarvan wordt de modelleringstijd die is gestoken in het modelleren van het bronsysteem hergebruikt door het Data Vault model direct van het geïmplementeerde brondatamodel af te leiden, zie afbeelding 1. In het geval van de industriële organisatie zal dit betekenen dat met behulp van transformatie snel en efficiënt een DWH-



Afbeelding 1: Transformatieproces.



Afbeelding 2: Vergelijking met conceptueel datamodel.

oplossing kan worden gerealiseerd voor het Asset Management vraagstuk. Op een eenvoudige wijze kan een eerste ERP-implementatie worden ontsloten naar het DWH, waarna deze kan worden vertaald naar de business met behulp van een rapportageomgeving. Vervolgens kunnen alle overige ERP-implementaties worden ontsloten en geïntegreerd, en ook de workflow-applicatie en het in eigen beheer ontwikkelde document management systeem kunnen worden toegevoegd. Wat blijkt; op een bepaalde Plant is Equipment ook gekoppeld aan Units, het ging hier om de heftrucks die niet aan een specifieke Tag toebehoren. In het document managementsysteem blijken ook juridische documenten opgenomen te zijn ten aanzien van de gehele Plant. Gelukkig is een Data Vault model zeer flexibel en kan het eenvoudig worden uitgebreid met Linktabellen om zo alle datastructuren te huisvesten. Maar hoe kan onze modelleur nu bepalen welk datamodel correct is en wat een afwijking? Wat is nu de waarheid en hoe vertaalt dat zich naar een rapportage?

Modelleren van een Data Vault

Door te kiezen voor een modelleringaanpak kunnen de vragen uit de voorgaande paragraaf worden afgevangen. In plaats van uit te gaan van een brondatamodel begint de modelleur met het opstellen van een conceptueel datamodel. Dat is een datamodel dat weergeeft hoe de business aankijkt tegen zijn eigen data. Vandaar dat dit modelleren wordt genoemd en niet transformeren. Dit conceptuele datamodel is de basis voor zowel het Data Vault model van het DWH als het dimensionale datamodel van de rapportageomgeving. Het conceptueel datamodel kan tevens worden gebruikt voor het toetsen van de datamodellen van de bronsystemen. Door het vergelijken van het conceptueel datamodel met het brondatamodel kan bijvoorbeeld worden gecontroleerd of de data in het ERP-systeem werkelijk opgeslagen zijn zoals het door de business bedoeld is. Eventuele verschillen tussen bronnen en het Data Vault model zijn nu verklaarbaar en gerelateerd aan het conceptuele datamodel. Alle datamodel ontwerpbeslissingen kunnen nu worden gemotiveerd en onderbouwd. De modelleur heeft grip op de situatie.

In het geval van de industriële organisatie zal door het vergelijken van de brondatamodellen met het conceptueel datamodel blijken dat Equipment wordt gekoppeld aan Units en dat ook nog juridische Documenten zijn gekoppeld op het niveau van Plant. Dit alles wijst op oneigenlijk gebruik van de Asset

Management toepassing. De business kan nu beslissen hoe om te gaan met deze issues. Het conceptuele datamodel kan worden aangepast, maar het is ook mogelijk om de bron op te schonen of het issue te accepteren. Het vinden van dergelijke issues kan worden gezien als de eerste resultaten die het DWH-project oplevert aan de business.

Voor- en nadelen van beide opties

Maar elk voordeel komt niet zonder zijn nadeel. In het geval van modellering zal het betekenen dat voor het ontwikkelen van het Data Vault model inclusief een conceptueel datamodel meer werk zal moeten worden verzet dan voor het uitvoeren van een transformatie binnen dezelfde project scope. Daar staat tegenover dat bij het transformeren naar een Data Vault model de modellering-issues onder het tapijt worden geschoven totdat deze bij de rapportages opduiken. Dit komt omdat transformatie een meer technische aangelegenheid is dan modellering, waarbij de focus is gericht op het realiseren van een systeem en niet op de issues die ontstaan bij het gebruik daarvan. De grote vraag is waar de organisatie behoefte aan heeft. Als de grootste drijfveer achter het DWH-project het snel beschikbaar krijgen van informatie is, dan kan transformatie de voorkeur krijgen. Maar als het creëren van een 'single point' van de waarheid de grootste drijfveer is, en de extra investering aan tijd en geld geen probleem, dan kan modellering de beste optie blijken te zijn. Dit alles betekent niet dat als eenmaal begonnen is met transformeren het niet meer mogelijk is om later alsnog een modelleeraanpak te starten, overstappen is altijd nog mogelijk.

Conclusie

Om tot een Data Vault model te komen is het mogelijk om tussen twee datamodelloopties te kiezen. Het is mogelijk om vanuit de business eisen een Data Vault model te modelleren door eerst een conceptueel datamodel op te stellen, gevolgd door een logisch en fysiek datamodel. Een andere optie is om een brondatamodel te transformeren naar een Logisch Data Vault model. Beide opties hebben hun voor- en nadelen waarbij de optie transformeren een meer praktisch, efficiënt karakter heeft en de optie modelleren zal leiden tot een betere definitie van de informatie. Uiteindelijk is het raadzaam om aan het begin van een project te bepalen welke optie de voorkeur geniet om tot een Data Vault te komen.

Ing. A.L. Pieters MSc is Senior Informatie Analist bij Logica.