

VAN DER LANS

Definities: Bij de loodgieter thuis lekt het



De laatste maanden ben ik intensief bezig geweest met het introduceren van een architectuur voor Business Intelligence systemen, genaamd het *Data Delivery Platform*. Als je een nieuw concept introduceert, moet je er uiteindelijk ook een definitie voor bedenken. Omdat ik zelf al een goed beeld had van wat het DDP inhield, leek mij dit een simpele opgave: even achter mijn laptop gaan zitten, wat friemelen met Word en binnen een uurtje is die definitie klaar. Dat viel uiteindelijk erg tegen. De hoofdreden dat dit een forse opgave bleek te zijn, was dat ik wilde dat de definitie van het DDP zou passen bij de bestaande verzameling van definities van andere datawarehouse gerelateerde concepten. Denk hierbij aan de definities van de concepten datawarehouse en Business Intelligence. Echter, er bestaan bijna geen algemeen geaccepteerde definities en als ze al bestaan, zijn ze niet erg eenduidig. En dit wordt heel duidelijk als je zelf zo'n definitie wilt bedenken.

Om te illustreren wat ik bedoel zal ik als voorbeeld de definitie van het begrip datawarehouse onder de loep nemen. Hierbij gebruik ik de populaire definitie van Bill Inmon: *A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process.*

Initieel lijkt deze definitie glashelder, maar bij nadere bestudering is zij dat niet. Bijvoorbeeld, hoe moeten we *collection of data* interpreteren? Betekent dit dat de gegevens in het datawarehouse fysiek als één verzameling opgeslagen moeten worden, of is het voldoende als de gegevens als één verzameling gepresenteerd worden? Op basis van de definitie zelf is dat niet vast te stellen. Echter, op basis van andere artikelen van Inmon kan je concluderen dat hij ervan uitgaat dat gegevens werkelijk opgeslagen worden. Kortom, deze definitie van datawarehouse is niet volledig.

In andere artikelen heeft Inmon opmerkingen gemaakt over wat wel of niet een datawarehouse is. Bijvoorbeeld in het artikel *What a Data Warehouse is Not* (bij BeyeNetwork.com gepubliceerd) geeft hij aan dat een zogenaamd active datawarehouse geen datawarehouse is. Deze term is door Teradata geïntroduceerd en zij hebben daar een bepaald idee bij. Het vreemde is echter dat wat Teradata hieronder verstaat wel degelijk voldoet aan Inmon's definitie. Inmon vindt echter van niet, omdat een datawarehouse de 'clerical community' niet hoort te ondersteunen. Maar waar staat dat dan in de definitie?

Ik heb nu als voorbeeld de definitie van datawarehouse genomen. Dat is niet omdat ik vind dat dit een slechte definitie is.

Ik heb deze gekozen omdat de term datawarehouse zo fundamenteel is dat we daar toch eigenlijk wel een volledige en eenduidige definitie van zouden moeten hebben. Maar helaas. We kunnen een vergelijkbaar teleurstellend betoog houden over de term Business Intelligence, waarvoor niet echt een algemeen geaccepteerde definitie bestaat. Hans Lamboo, hoofdredacteur van DB/M, heeft recent een column over deze term geschreven. Hierop kwamen direct diverse positieve en negatieve reacties. Niet geheel verwonderlijk, want als er geen definitie is die algemeen geaccepteerd is ontstaan er discussies.

Er bestaan wel definities van BI. De website van SearchData-management geeft bijvoorbeeld de volgende: *BI is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions.* Maar ik heb ook deze definitie gevonden: *Business Intelligence is a popularized, umbrella term used to describe a set of concepts and methods to improve business decision making by using fact-based support systems.* En dit zijn toch echt verschillende definities. Als u gaat zoeken op het internet, zult u er nog tientallen andere vinden.

Kortom, we gebruiken allemaal te pas en te onpas termen en doen net alsof iedereen daar hetzelfde onder verstaat. Maar dat kan alleen als alle termen eenduidig gedefinieerd zijn en als de definities algemeen geaccepteerd zijn, maar dat zijn ze absoluut niet.

Het is wel vreemd om te constateren dat als wij voor een klant een datawarehouse-omgeving moeten optuigen, wij de eersten zijn die komen vertellen hoe belangrijk het is dat we haarscherpe definities van alle data-elementen hebben. We moeten bijvoorbeeld precies weten wat nu wel en wat nu niet een klant is, en wat nu precies een vlucht is. Daarbij vinden we dat deze definities organisatiebreed geaccepteerd dienen te worden.

Maar zelf hebben we bijna geen enkele definitie. Twee gezegdes komen dan bij mij op. Ten eerste, "de pot verwijt de ketel dat 'ie zwart ziet" en "bij de loodgieter thuis lekt het".

Soms heb ik het idee dat er nieuwe termen geïntroduceerd worden die in feite gewoon synoniemen zijn voor iets dat we al hadden. Als we van de oude en de nieuwe term definities zouden hebben, kunnen we aangeven dat het gewoonweg een hippere term betreft. Dus laten we het voorbeeld geven en onze termen eens goed definiëren. Zorg dat wij de loodgieters worden waar het thuis *niet* lekt.

Rick van der Lans is zelfstandig IT-consultant.