

amBivalent

# All quiet on the datawarehousing front?

Tijdens het Datawarehousing en BI Congres op 2 oktober zijn de resultaten van een tweetal onderzoeken gepresenteerd: het Nationaal Datawarehouse Onderzoek (DWO 2008) van Atos Origin/Array en de BI Survey van Nigel Pendse. Beide onderzoeken hadden een hoog 'er is niet zoveel gebeurd'-gehalte. Maar is dat ook zo?

Uit het Nationaal Datawarehouse Onderzoek 2008 blijkt dat de volwassenheid van datawarehouses en BI in Nederland zich nagenoeg op hetzelfde niveau bevindt als in 2005. Zo'n conclusie leidt nu niet meteen tot een euforische stemming. De eerste gedachte is dat we met z'n allen drie jaar lang hebben stilgestaan. Meer een nationale *surplace* of het grote watertrappelen. Volgens Frank Buytendijk is er iets anders aan de hand. Zijn argumentatie is (samengevat) dat in de afgelopen drie jaar de (BI-) wereld veel complexer is geworden en dat het dus juist een prestatie is dat we dit niveau hebben weten vast te houden. "We moeten er juist trots op zijn". En inderdaad, dit zou wel eens waar kunnen zijn! In ieder geval is dat een leukere conclusie dan dat we de afgelopen drie jaar zijn ingedut. Een andere argumentatie die ik kan bedenken sluit aan bij de rol die de zogenaamde 'champions' spelen in de eerste volwassenheidsfasen. Kort gezegd: de bereikte volwassenheid berust voor een groot deel op specifieke personen ('champions'). "Voor ingewikkelde rapportages moet je bij Peter zijn". Dat is ook de kwetsbaarheid van de situatie. Schuiven deze champions op (of verlaten ze de organisatie) dan kan dat een gevoelige deuk geven in de volwassenheid van BI en datawarehousing. Kennis en ervaring verdwijnen *down the drain*. Een periode van drie jaar tussen het huidige (2008) en vorige onderzoek (2005) maakt het aannemelijk dat ook zo'n ontwikkeling speelt. Waaraan ligt de identieke score: massale inactie; toegenomen complexiteit; of de gevolgen van organisatorische doorstroming?

Ook de BI Survey van Pendse suggereert weinig verandering. Zo bleek uit de resultaten dat de hoeveelheid data waarmee we werken niet is toegenomen. *I find that hard to believe*. Als wordt bedoeld dat we nog steeds naar dezelfde hoeveelheid data kijken is dat nogal flauw omdat er natuurlijk fysieke grenzen zitten aan de hoeveelheid data die een kubus kan bevatten. Maar dat bedoelt Pendse niet. Op de vraag "are data volumes really exploding?" is zijn antwoord: "volumes have changed little and are surprisingly small". Is er in navolging van de kredietcrisis nu ook een datacrisis?

Hoe meer data je bewaart, hoe minder je overhoudt? Uit het DWO 2008 blijkt dat bijna 1 op de 5 organisaties een datawarehouse heeft van minder dan 1 GB. Voordat we nu wereldwijd onze datawarehouseservers opdoeken en in plaats daarvan memory sticks gebruiken (wat wel de suggestie lijkt) zou ik een alternatieve kijk op deze cijfers willen aanbieden. Pendse heeft het over mediaanwaardes en over 'typical applications'. Nu is een applicatie niet hetzelfde als een datawarehouse. En ga ik ervan uit dat meerdere applicaties gebruik maken van de data die zijn opgeslagen in een datawarehouse. Andersom gezegd: een specifieke applicatie gebruikt slechts een deel van de datawarehousedata. Je mag je dus afvragen of je door te kijken naar een applicatie een juist beeld krijgt van de omvang van het totale datawarehouse. Het komt toch neer op het beschrijven van een huis door via de brievenbus naar binnen te kijken. Maar goed, laten we er even vanuit gaan dat dit goed is ingeschat. Dan zitten we nog met de mediaanwaarde. Eenvoudig gezegd is dit de middelste waarde in een reeks geordende getallen. Stel dat we de omvang weten van vijf datawarehouses: 1-1-5-100-1000 GB. De mediaanwaarde is hier 5 GB. Doen bij een volgend onderzoek negen bedrijven mee dan kan de reeks er als volgt uitzien: 3-4-4-4-5-200-1000-1500-2000. U snapt het al, de mediaanwaarde is nog steeds 5 GB. Vragen die je aan jezelf kunt stellen: zou je op basis van de eerste reeks cijfers constateren dat datawarehouses 'surprisingly small' zijn? Zou je op basis van de tweede reeks van cijfers constateren dat er niet veel veranderd is? Let wel, dit zijn slechts voorbeeldreeksen en niet de echte volumes. Maar het zou best kunnen dat de vele nieuwe, relatief kleine datawarehouses de reeds langer bestaande oudere datawarehouses van inmiddels respectabele omvang 'compenseren'. Waardoor de mediaanwaarde constant blijft. Want dat er reeds vele Terabyte sized datawarehouses zijn wordt aangetoond door weer andere onderzoeken. Een kwart van de organisaties die deelnamen aan het DWO 2008 meldt meer dan 2 TB aan data te hebben. Drie jaar eerder was dat 14 procent. Misschien dat er toch iets meer aan de hand is dan Pendse ons wil doen geloven.

Tenslotte nog dit: volgens Pendse (en anderen) is SAP BW al jarenlang hekkensluiter in wat nu de Business Benefits Index heet. Kan iemand hiervoor alsjeblijft een alternatieve interpretatie bedenken? Bij mij stekt het nu even.

**Paul van der Linden** (Paul.vanderLinden@nl.ibm.com) is senior managing consultant BI/IOD bij IBM Global Business Services.