

VAN DER LANS

Over datawarehouses en rioleringen



Thuis hebben wij, net als veel andere gezinnen, een riolering. Hij zat er in toen we het huis kochten. Op zich wel prettig om een riolering te hebben, want anders wordt het allemaal wat onfris. Maar we hadden er niet om gevraagd. Wat we wilden was een systeem dat de onprettige zaken zou afvoeren op het moment dat we het toilet zouden verlaten. Hoe dat systeem zou functioneren was voor ons niet belangrijk. Als dit afvoersysteem het onaangename de ruimte in zou schieten, of als het op een magische manier het afval zou oplossen, zouden wij thuis het ook goed gevonden hebben. Maar het bleek uiteindelijk een riolering te zijn bestaande uit een complexe infrastructuur opgebouwd uit allerlei buizen onder de grond die mijn huis verbindt met rivieren en oceanen.

Eigenlijk geldt hetzelfde voor een datawarehouse-omgeving. Niemand heeft daar ooit om gevraagd, zeker de gebruikers niet. Wat zij wilden was 'iets' om hun beslissingsproces te verbeteren. Hiervoor zijn waarschijnlijk rapportage- en analyse-producten nodig die gegevens in de goede vorm presenteren en de functies bieden die de gebruikers nodig hebben. Dit was hun eigenlijke vraag. Ik heb nog nooit een gebruiker uit de business ontmoet die vroeg of we een grote bak met gegevens voor hem wilden opbouwen. Een datawarehouse-omgeving is dus op een bepaalde manier te vergelijken met een riolering. Trouwens, rioleringen en warehouses hebben meer zaken gemeen, zo hanteren we bij beide het gezegde *Garbage In Garbage Out*, en voor beide geldt dat als we er teveel doorheen pompen er verstoppingen optreden.

Natuurlijk hebben de meeste rapportageproducten, om te kunnen functioneren, toegang tot gegevens nodig. Maar moet dat dan per se een datawarehouse of datamart zijn? Nee, niet echt. Technisch gezien zijn er vele oplossingen mogelijk. Misschien is het draaien van de producten op de productiesystemen wel een optie?

Jaren geleden zijn er diverse Business Intelligence architecturen bedacht om de rapportageproducten van gegevens te voorzien, zoals Kimball's bus-architectuur en Inmon's Corporate Information Factory. Deze architecturen zijn bedacht in een tijdperk waarin de hardware en de software er toch wat anders uitzagen dan nu. Toen waren er nog geen datawarehouse appliances, bestond solid state disk niet, hadden we geen producten zoals QlikView en Microsoft's PowerPivot, waarmee we miljoenen rijen met gegevens in het geheugen van een laptop kunnen laden, en over Cloud werd zeker nog niet gesproken.

Technieken, architecturen en methoden hebben een houdbaarheidsdatum. Op een gegeven moment zijn ze niet meer geschikt. Bijvoorbeeld, de techniek die ooit gebruikt werd om met de voet het spinnewiel te draaien, waarover veel geschreven is en enorme discussies zijn gevoerd, is haar houdbaarheidsdatum voorbij. Of de architectuur die op de prairie geadviseerd werd voor het ontwikkelen van stevige, windvaste en droge indianenwigwams, is ook niet meer van levensbelang. Met de komst van nieuwe technologieën kan het zijn dat een techniek, architectuur of patroon vervalt.

Met al die nieuwe technologieën die de afgelopen jaren verschenen zijn, wordt het tijd dat we ook eens gaan controleren of de houdbaarheidsdatum van sommige van onze technieken en architecturen reeds gepasseerd is. Wat zeker geen zin heeft is om blind de oude technieken en architecturen te blijven gebruiken wanneer nieuwe technologie ingezet wordt. Als we dit toch doen, halen we meestal niet het maximale uit die nieuwe technologie. Misschien heeft het, bijvoorbeeld, geen zin meer om datamarts te bouwen als extensies op een centraal datawarehouse. Misschien zijn de nieuwe rapportageproducten al zo krachtig dat bij het opstarten alle benodigde gegevens ter plekke in het geheugen geplaatst kunnen worden. Waarschijnlijk is dan nog wel een datawarehouse nodig.

De gebruikers hebben nooit om een datawarehouse gevraagd. Zij zullen dan ook geen emotionele problemen hebben als de datawarehouse-omgeving verandert of zelfs geheel verdwijnt. Wij hebben ze ooit moeten ontwikkelen vanwege allerlei technische redenen, maar de vraag is of die redenen nog steeds valide zijn. Laten we niet te krampachtig aan bestaande oplossingen, architecturen en technieken blijven hangen. Soms moeten we ook even op het pak melk kijken welke datum er op staat. Het wordt tijd dat we die datum voor de architecturen en technieken die we de afgelopen twintig jaar succesvol gebruikt hebben, ook gaan bestuderen.

Terugkomend op de riolering, misschien ontwerpen ze ooit een nieuwe technologie om het onappetijtelijke te laten verdwijnen. Iets waarvoor zo'n complexe infrastructuur niet nodig is. Ik, en ik denk velen met mij, zouden het prima vinden; ook geen verstoppingen meer. De nieuwe technologie voor datawarehouse-omgevingen is al gearriveerd, nu kijken hoe wij die het beste kunnen inzetten.

Rick van der Lans is zelfstandig IT-consultant.