

VAN DER LANS

Het datawarehouse geen database meer?



Iedereen hanteert zo zijn eigen favoriete definitie van de term Business Intelligence. Wanneer we naar de verschillende definities kijken, dan komen de meeste meestal op het volgende neer: Business Intelligence is het geheel van vaardigheden, methoden, processen, technologieën en applicaties ter ondersteuning van het beslissingsproces van een organisatie. Met andere woorden, het gaat om het ondersteunen en (hopelijk) verbeteren van de beslissingsprocessen.

In de meeste gevallen wordt hiervoor een groot datawarehouse opgebouwd. Alhoewel, in waarschijnlijk geen enkele organisatie bestaat de datawarehouse-omgeving werkelijk uit slechts één grote database gevuld met brongegevens en waarop alle rapporten en analyses draaien. Meestal bestaat zo'n omgeving uit een aantal databases, zoals het ODS, het centrale datawarehouse en diverse datamarts. Wel kunnen we stellen dat nagenoeg alle rapporten en analyses gevoed worden met gegevens die in één van de databases van de datawarehouse-omgeving opgeslagen zijn. Maar hoelang kunnen we dit nog volhouden? Kunnen we wel alle vormen van het beslissingsproces met een dergelijke databasegebaseerde oplossing ondersteunen? We lichten twee voorbeelden toe waarin dit onmogelijk bleek te zijn.

In een kipverwerkingsbedrijf in Amerika worden tienduizenden kilo's kip per dag verwerkt. De prijs van de meeste kipproducten is afhankelijk van het gewicht. Wat bij het verwerkingsproces belangrijk is, is de temperatuur. Hoe heter, hoe meer vocht in de kip verdampt en hoe kouder, hoe meer vocht er blijft zitten. Hoe meer vocht in de kip blijft zitten, hoe minder kip per cm^3 nodig is om een bepaald gewicht te halen. Kortom, warmer betekent minder water en is er dus meer kip nodig en hiermee worden de kosten voor het bedrijf hoger. Om dit proces goed in de gaten te houden zijn dashboards ontwikkeld waarin de temperatuur van het bereidingsproces continu in de gaten gehouden kan worden en waarmee de hoeveelheid vocht in het kippenvlees continu bestudeerd kan worden.

Eén graad teveel of te weinig heeft grote financiële consequenties. De data waarmee de dashboards gevoed worden, komen niet uit een datawarehouse, maar direct uit de operationele systemen. Het eerst kopiëren naar een datawarehouse en vervolgens naar een datamart zou teveel vertraging opleveren. Eén minuut te laat reageren kan veel geld kosten.

Opmerking: dit voorbeeld is ontleend aan een presentatie verzorgd door LogiXML.

Een ander voorbeeld is van een internationaal opererende organisatie van elektronische apparatuur die continu alle gepubliceerde tweets waarin de naam van de organisatie voorkomt wil scannen en analyseren. Dit kan door middel van semantische en sentimentanalyse (wat momenteel door diverse producten ondersteund wordt). Als er een negatieve tweet voorbijkomt, moet dit aan een operator gemeld worden en deze moet vervolgens beslissen of er al dan niet gereageerd zal worden.

Wederom, dit beslissingsproces zou te traag werken als eerst alle tweets in een datawarehouse verzameld moeten worden en ze daarna naar een datamart doorgekopieerd moeten worden voordat de gegevens voor analyse beschikbaar zijn. Het is beter deze externe bron direct te benaderen en direct te analyseren.

En zo zijn er al veel voorbeelden in het bedrijfsleven van beslissingsproces ondersteunende systemen waarbij het eerst opslaan van de gegevens in een database niet meer de gekozen oplossing is. Zeker in die omgevingen waarbij snel beslissingen genomen moeten worden, dus waarbij de tijdsduur tussen het moment waarop een actie plaatsvindt en het moment waarop er door middel van een beslissing op gereageerd moet worden, kort tot zeer kort is.

Omdat veel Business Intelligence omgevingen slechts veredelde rapportageomgevingen ofwel verslagleggingomgevingen zijn, is het te begrijpen dat hier meestal voor een databasegebaseerde oplossing gekozen wordt. Maar kijken we naar het totale beslissingsproces van een organisatie, dan moeten we concluderen dat ter ondersteuning van veel beslissingen een database niet de ideale technische oplossing is. Voor de ene vorm van Business Intelligence is uiteraard een databasegebaseerde oplossing het beste alternatief, maar voor een andere zal bijvoorbeeld replicatie of datavirtualisatie nodig zijn. Uiteraard zal het dan wel een uitdaging worden om alle gegevens consistent te houden.

Het doel van Business Intelligence was en is het ondersteunen van het beslissingsproces en niet het opbouwen van databases. Laten we dus uit dit keurslijf stappen en ervan uitgaan dat een databasegebaseerde oplossing slechts een technische oplossing is; één van de alternatieven.

Rick van der Lans is zelfstandig IT-consultant.