



## Vooraankondiging van een open source datawarehouse management system

In het verleden heeft Nippur via verschillende media gewag gemaakt van het feit dat het ontwikkelen van datawarehouses en bijbehorende laadprocessen een tijdrovende en kostbare aangelegenheid is. Veroorzaakt door het handmatig moeten ontwikkelen van datamodellen en bijbehorende laadscripts.

Herhaaldelijk is hierbij gepleit voor het inzetten van zogenaamde datawarehouse management systemen. Als consultancy onderneming is Nippur al jaren gewend te werken met die paar producten die daartoe ontwikkeld zijn. We kennen dan ook de voordelen die deze hulpmiddelen bieden.

Het is dan wrang om te moeten constateren dat de keuze in technologie beperkt is. Waar al oplossingen bestaan zijn deze buitengewoon kostbaar in aanschaf en dus lijken voorbehouden aan de happy few: de grotere multinationale organisaties met grote BI competence centers.

Dankzij de komst van het Data Vault modelleringconcept is een interessante nieuwe situatie ontstaan. Dit concept is zo generiek van opzet dat het zich leent voor automatisering. Dat weten wij als geen ander omdat wij als Data Vault specialist al vele klanten hebben geholpen met de ontwikkeling van Data Vault georiënteerd datawarehouses. Telkens ontwikkelden wij hulpmiddelen (scriptgenerators) om de repetitieve (en dus saaie) handelingen te automatiseren. Wij zagen de voordelen hiervan omdat we enorme versnellingen in de implementatie van een datawarehouse konden bewerkstelligen. In een enkel geval tot wel 80%(!) sneller dan het traditioneel handmatig bouwen van een BI architectuur. Met hetzelfde budget kunnen klanten dan meer vraagstukken onderzoeken en implementeren. En als het aan ons ligt is dát het echte leuke werk. Het begrijpen van de klant, het meedenken hoe een business vraagstuk op te pakken zodat de klant echte goede business cases kan definiëren en oppakken.

Dit gegeven én het eerder genoemde feit dat bestaande management systemen onbetaalbaar zijn heeft ons doen besluiten een datawarehouse management systeem te ontwikkelen dat wij in open source ter beschikking willen stellen aan iedereen met een data integratie behoefte. Dat project en tevens product noemen we Quipu. Naar analogie met de quipu's die archeologen vonden in en rond Cuzco (ook wel geschreven als QOSQO) in Peru.

Khipu's of quipu's waren bij de Inca's een soort koorden die gebruikt werden om getallen weer te geven. Dit deden ze door middel van knopen. Er werden meerdere koorden aan een koord geregen. Het tellen van getallen gebeurde door het leggen van knopen, waarbij het decimale stelsel werd toegepast. De onderste knopen gaven de eenheden aan, daarboven de tientallen. De soort van het te tellen eenheden werd door de kleur aangegeven. [bron: wikipedia].



Figuur 1. Voorbeeld van een Quipu.

Binnenkort ziet versie 1.0 van Quipu het licht. Vandaar deze vooraankondiging.

## Quipu

In augustus 2008 is QOSQO opgericht om een tweetal zaken op te pakken. Enerzijds het bieden van datawarehouse services (lees: 2nd line support op Data Vault georiënteerde datawarehouses veelal ontwikkeld door zusterbedrijf Nippur). Anderzijds de ontwikkeling en lancering van Quipu, het open source datawarehouse management systeem. We kennen al enige tijd open source BI toepassingen die functionaliteit bieden op het vlak van rapportage, ETL, databases, mining, wiki's etcetera. De data integratie component wordt echter door geen van deze spelers ingevuld. Vermoedelijk vanwege de zeer specifieke kennis die nodig is om Data Vault te kunnen doorgronden, Data Vault storage modellen te kunnen afleiden en de bijbehorende ETL te kunnen genereren.

Binnenkort lanceren we versie 1.0 van Quipu. Hiermee kan een brongeoriënteerd datawarehouse worden gecreëerd. De gebruiker wordt daarin grafisch via een aantal logische stappen begeleid in het proces van het ontwerpen, creëren en beheren van een technisch robuust datawarehouse.

In eerste instantie kan een bronmodel worden ingelezen en opgeslagen in de Quipu repository. Vervolgens zorgt interne algoritmieken door toepassing van een aantal afleidingsregels voor de opstelling van een Data Vault datamodel. Handmatig kunnen vervolgens wijzigingen worden doorgevoerd, waarna het aangepaste datamodel wederom in de repository wordt opgeslagen. Vervolgens worden standaard mappings tussen bron en target model gegenereerd die door de architect kunnen worden aangepast en aangevuld. Na akkoord genereert Quipu het datawarehouse datamodel en de bijbehorende laadscripts.

Er ligt een roadmap klaar voor toekomstige releases van Quipu waarbij data mart generatie (en bijbehorende ETL) wordt gerealiseerd. Tevens zijn er ideeën hoe om te gaan met de big 'T', of grote transformatieslag die op enig moment onvermijdelijk is om meer complexe business logica toe te voegen. Er zal geen ETL tool worden ontwikkeld. Het ligt wél voor de hand de metadata van de transformatieprocessen van belangrijkste ETL leveranciers in de repository op te nemen zodat gerapporteerd kan worden over de hele datawarehouse procesflow. Hiermee wordt voorzien in de toenemende behoefte aan inzicht in lineage en aan de eisen gesteld op het gebied van auditability en traceability.

Wij hebben de overtuiging dat deze eerste release al vele ondernemingen zal helpen bij het sneller neerzetten van een integratieplatform. De eerste toepassingen hebben zich bij klanten al bewezen.

## About Data Vault

The Data Vault Model is a flexible, scalable, and consistent "common foundational warehouse model architecture", invented by Daniel Linstedt. The modeling technique is covered by an implementation methodology. The Data Vault methodology is a set of data warehousing implementation components, consisting of project plans, design constructs, ETL loading templates, risk analysis, and repeatable, consistent build out strategies. The project / implementation methodology has its basis in SEI/CMMI Software engineering Level 5 components. More info can be found on [www.danlinstedt.com](http://www.danlinstedt.com) and [www.datavaultinstitute.com](http://www.datavaultinstitute.com). Interested in getting trained in the Data Vault methodology and modelling technique? [www.datavaultacademy.com](http://www.datavaultacademy.com)

**QOSQO**  
datawarehouse services

Karel Doormanlaan 1b, 5688 BP Oirschot (NL)  
Telefoon: +31 (0)499 577 562, Fax: +31 (0)499 577 059  
E-mail: [info@qosqo.nl](mailto:info@qosqo.nl), <http://www.qosqo.nl>  
Bank: ABN-AMRO 42.16.43.617  
K.v.K. Midden Brabant: 17226645, BTW 8194.65.847.B01