

Nederlandse 'eXtreme ETL Programming'-technologie

# Hoofdrol voor metadata bij bedrijfsanalyses

Robbert Hoeffnagel

**Hoe kan een specialist in het werven en selecteren van hooggekwalificeerd personeel het beste zijn omzet vergroten? Is het dan beter om meer met klanten te bellen of is het juist verstandiger om meer bezoeken af te leggen?**

Om grip op dit soort kwesties te krijgen, heeft Michael Page International in Amsterdam een analytische toepassing van Dynalytical Solutions in gebruik genomen. "Natuurlijk kennen we onze business wel, maar nu weten we precies aan welke knoppen wij moeten draaien om het nog beter te doen. Bovendien kunnen we nu meer en betere relaties tussen voorheen los van elkaar staande gegevens leggen", vertelt Anton den Bak, Finance Director van Michael Page International. Hij is financieel verantwoordelijk voor de activiteiten van het bedrijf in een aantal Noord-Europese landen, waaronder Nederland.

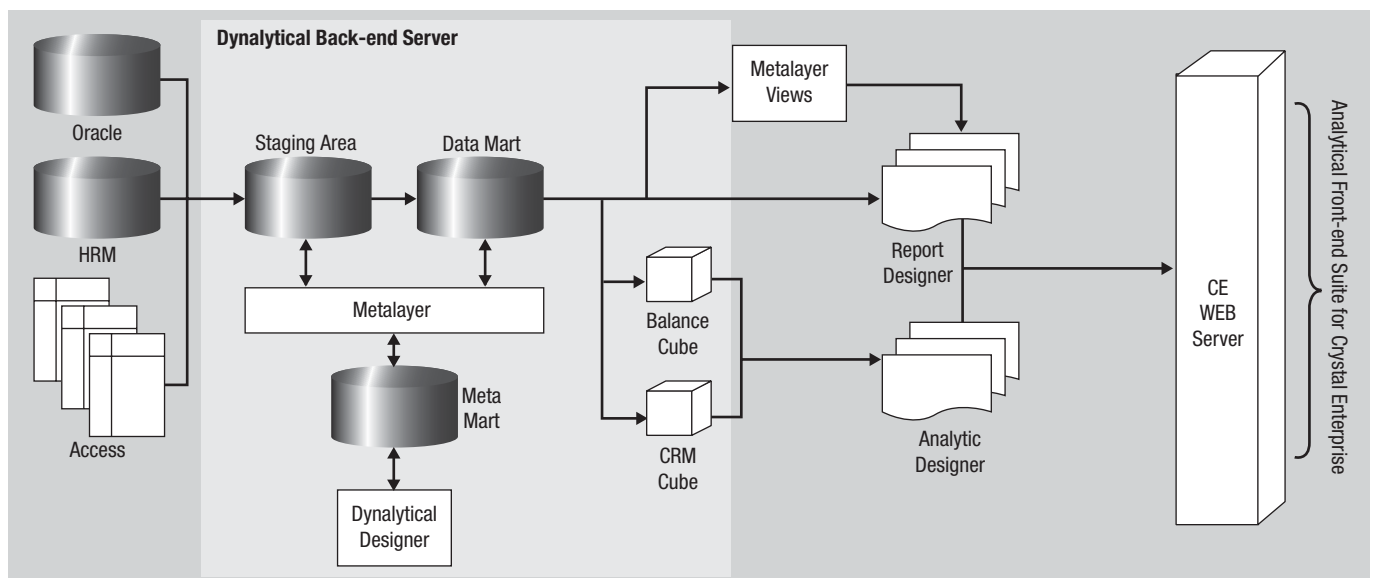
## Resultaat

Michael Page houdt zich bezig met het werven en selecteren van personeel op het gebied van finance, verkoop en marketing, juridische zaken en belastingen, ICT, engineering en supply chain, evenals human resources. "Wij maken onder andere gebruik van een database waarin wij van veel mensen de

belangrijkste kenmerken vastleggen." Naast een drukbezochte website met ruim 75.000 unieke bezoekers per maand, werkt het bedrijf met consultants die de contacten met de werkgevers onderhouden.

Den Bak vervolgt: "U kunt zich voorstellen dat er nogal wat informatie in deze organisatie rond gaat. Zo kennen wij naast de database met potentiële kandidaten, ook een groot aantal mensen die zich via e-mail geabonneerd hebben op ons vacatureoverzicht. Ook beschikken we over een uitgebreid overzicht van bedrijven die op zoek zijn naar medewerkers. Daarnaast hebben wij consultants in dienst die veelal gespecialiseerd zijn in bepaalde markten of sectoren. Hun prestaties willen wij uiteraard kunnen meten. En niet te vergeten dienen de resultaten van al deze inspanningen op een goed gestructureerde manier te worden vastgelegd in een financieel systeem.

Alles hangt echter met alles samen. Om een voorbeeld te noemen: het aantal contacten dat een consultant heeft met zijn klanten, heeft natuurlijk invloed op het financiële resultaat dat hij of zij behaalt. Maar hoe zit die relatie nu precies in elkaar? De ene consultant is constant op pad en bezoekt dagelijks klanten. Een ander blijkt veel meer telefonisch te werk te gaan. Wat scoort beter? En hoe komt dat nu precies? Bovendien hebben wij te maken met veelal zeer zelfstandig opererende medewerkers.



Afbeelding 1: De architectuur van Dynalytical Solutions.

Dit wil zeggen dat wij niet als een schoolmeester hun activiteiten gaan controleren. Toch willen we echter op basis van goede informatie met elkaar kunnen praten. Een consultant kan wel zeggen dat 'het lekker gaat', maar dat moet dan ook wel uit de administratie blijken. Wat zouden we bovendien kunnen doen om het nog beter te laten gaan?"

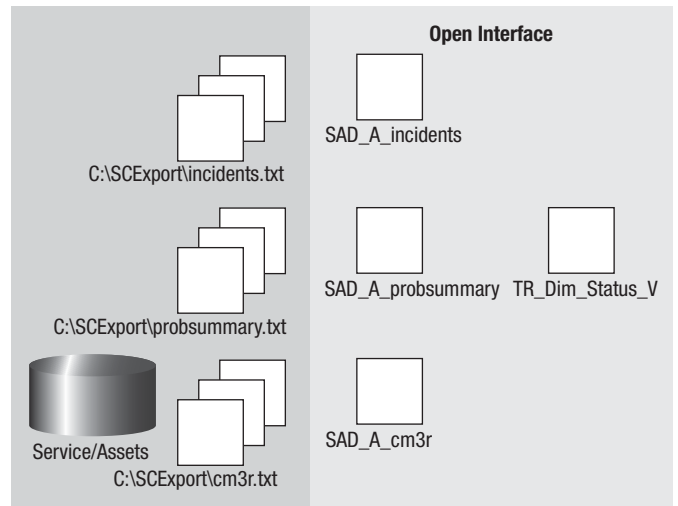
## Tal van relaties

Binnen Michael Page International bestond voorheen geen proces dat al dit soort gegevens op een gestructureerde manier bij elkaar brengt. Er werd weliswaar gewerkt met spreadsheets waarin tal van gegevens werden vastgelegd, maar het op een dusdanige manier koppelen van de informatie in de diverse werkbladen dat hier zinvolle conclusies uit konden worden getrokken bleek niet eenvoudig. "Bovendien is de hoeveelheid data waarmee we werken en het aantal bronsystemen waaruit deze gegevens afkomstig zijn te groot om op een zinvolle wijze met spreadsheets aan de slag te kunnen gaan. Wat wij in feite zochten was een Business Intelligence-oplossing die niet alleen alle gegevens uit de diverse bronsystemen bij elkaar bracht, maar die het bovendien mogelijk maakte om vervolgens analyses op al deze data los te laten. Die zijn er natuurlijk volop, maar wat ons voor ogen stond was een modelmatige manier van rapporteren waarbij wij zelf op eenvoudige wijze het te gebruiken model kunnen opbouwen, onderhouden en waar nodig aanpassen. Wij willen aantallen telefoontjes kunnen afzetten tegen bijvoorbeeld het aantal plaatsingen bij klanten. Maar ook willen wij inzicht hebben in de vraag wat bijvoorbeeld de doorlooptijden van eerste contact tot feitelijke plaatsing van kandidaten zijn, hoeveel reacties er op advertenties komen, en in welke relatie dit staat ten opzichte van plaatsingen of doorlooptijden enzovoort. Op deze manier willen wij – indien nodig – van een redelijk high-level 'management view' helemaal kunnen afzakken tot op operationeel niveau. Eigenlijk draait het voor ons om een hele simpele kwestie: in welke mate moeten wij aan welke knoppen draaien om het nog beter te doen in de markt?"

Stefan Reinerink, IT-manager van Michael Page International, kwam in aanraking met Dynalytical Solutions. Deze in Nieuwegein gevestigde firma houdt zich sinds januari 2004 bezig met het ontwikkelen van analytische applicaties. Het bedrijf gebruikt hierbij de zogeheten Dynalytical Toolkit. De tools die deel uitmaken van dit product dienen enerzijds voor het ontwikkelen en aanbieden van een aantal standaard producten als het voor service- en asset management bedoelde pakket SAMIS, maar kunnen anderzijds ook worden toegepast om maatwerkapplicaties te vervaardigen.

## Back-end server

Een hoofdrol is weggelegd voor een zogeheten 'back-end server'. Deze 'omgeving' wordt geïnstalleerd op een fysieke server en omvat een reeks van componenten die nodig zijn om data via ETL (extract, transform and load) processen beschikbaar te maken voor



**Afbeelding 2:** Voorbeeld van het laadproces van drie tekstbestanden.

een 'front-end applicatie' waarmee de feitelijke analyses worden gedaan. Hiertoe kent de back-end omgeving drie databases – een 'staging area', een 'datamart' en een zogeheten 'metamart' – en meerdere OLAP-cubes. Alles is gebaseerd op Microsoft-technologie: SQL Server en MS OLAP cubes.

"De gehele aanpak draait om het gebruik van metadata", zegt Freek Kamst, chief technical officer van Dynalytical Solutions. "Met de hulp van een aantal tools worden de formules en parameters die betrekking hebben op de ETL-processen, de datamodellen en de indexstrategie in een metadata-omgeving vastgelegd. Tijdens het ETL-proces worden deze regels vervolgens uitgelezen en wordt de omgeving direct aangepast. Dat heeft als voordeel dat wijzigingen in de informatiebehoefte snel kunnen worden doorgevoerd – het hoeft immers alleen in de metadata te worden aangepast – terwijl het resultaat daarvan bovendien direct online zichtbaar wordt.

Doordat dit allemaal door middel van metadata gebeurt, kan een forse tijdswinst worden geboekt ten aanzien van ontwikkeling en onderhoud", meent Kamst. "De tijdswinst die hierdoor ontstaat ten opzichte van het onderhouden van een traditionele BI-omgeving kan oplopen tot wel 80 procent. Binnen traditionele omgevingen gaat namelijk iedere keer als de informatiebehoefte verandert, veel tijd verloren met het aanbrengen van wijzigingen in de omgeving zelf. Een groot probleem hierbij is dat de tijd die gaat zitten tussen het moment waarop een wijziging nodig is en het moment waarop deze daadwerkelijk is doorgevoerd veel te lang is."

## eXtreme ETL Programming

Juist doordat de snelheid waarmee veranderingen op organisaties afkomen steeds hoger wordt, is het van groot belang dat de tijd die nodig is voor het implementeren van wijzigingen zo kort mogelijk is. Kamst hanteert hier – vrij naar de bekende programmeertechniek – een interessante benaming voor: eXtreme ETL Programming. "Met name het programmeren van complexe ETL-processen is een tijdrovende aangelegenheid", zegt Kamst.

“Daarom hebben wij op basis van de beginselen en concepten van eXtreme Programming het begrip ‘eXtreme ETL Programming’ gelanceerd. Hierbij speelt onze ‘Dynalectical Toolkit’ een hoofdrol.” Deze toolkit bestaat uit een framework, een ETL-library en een metadata-omgeving, waarmee een analytische applicatie kan worden gebouwd die wordt bestuurd via metadata. Deze applicatie is maatwerk voor de informatiebehoefte van de eindgebruikers in een bepaald domein, functiegebied of bedrijfsproces. “Bij de traditionele manier van het ontwerpen en ontwikkelen van een BI-omgeving worden eventuele problemen met betrekking tot de kwaliteit van gegevens en business-definities pas zichtbaar bij het testen van de reeds ontwikkelde ETL-componenten. Meestal is men dan al voor een langere periode bezig met het project en krijgt de projectleider te maken met allerlei niet voorziene activiteiten, om van de extra kosten en de overschrijdingen van deadlines nog maar te zwijgen.”

Wordt eXtreme ETL Programming toegepast, dan werkt het anders, vertelt Kamst. “Wijzigingen in de informatiebehoefte worden nu via een grafische gebruikersinterface doorgevoerd. Dat doen we met een tool dat wij de Dynalectical Designer noemen. Met dit hulpmiddel kunnen wijzigingen in het metadata-model worden aangebracht. Hierop wordt automatisch het datamodel in de zogeheten ‘staging area’ en de ‘datamart’ aangepast. Ook worden de benodigde ETL-processen gecreëerd en wordt de indexering in de staging area en de datamart geoptimaliseerd. Eventuele conversies of migraties worden eveneens automatisch via hulptabellen uitgevoerd. Het verversen van de OLAP-kubussen wordt tijdens het volgende laadproces uitgevoerd. Hierdoor worden eventuele problemen ten aanzien van business-definities of de kwaliteit van data snel zichtbaar.”

## Open interface

Daarmee zijn we ook direct aangekomen op de diverse stappen die in de back-end server worden doorlopen. Dit is weergegeven in afbeelding 1. “Via een open interface zijn wij in staat data te importeren uit tal van bronsystemen. We ondersteunen een hele reeks van formaten, waaronder OLEDB, ODBC, XML, Oracle, ascii en Excel.” Afbeelding 2 toont het laden van een tekstbestand, in dit geval worden drie tekstbestanden ingelezen. Kamst spreekt van een ‘high speed staging area’. “Deze is bedoeld voor het in bulk verwerken van data. Om de kwaliteit van de data te waarborgen, worden vijf stadia doorlopen, aangestuurd via de metadata die in de metamart zijn vastgelegd. Uiteindelijk komen de data in het datawarehouse terecht. Merk hierbij op dat er in het datawarehouse zelf géén verwerking van gegevens plaatsvindt. Wel wordt een volledige historie van alle stappen en stadia bijgehouden en vastgelegd. Het datawarehouse is bovendien altijd beschikbaar.”

Oorspronkelijk was het de bedoeling dat de eindgebruiker vervolgens met Crystal Reports tot de feitelijke rapportage kwam. De ondersteuning aan deze kant van het product is echter verbreed. Kamst: “We zijn aan de front-end nu productonafhankelijk.

De eindgebruiker kan dus net zo goed met bijvoorbeeld ReportNet van Cognos of Microsoft Reporting Services aan de gang gaan. Bovendien gaat het om een web enabled architectuur, zodat rapporten ook via een browser kunnen worden bekeken. Daarnaast is sprake van standaard ‘row level security’, wat met name voor het beveiligen van informatie belangrijk is.”

## Mogelijkheden

Wat zijn nu precies de voordelen van deze aanpak? Kamst en Reinerink zijn het er over eens dat de dimensionele structuur de kwaliteit van gegevens goed zichtbaar heeft gemaakt. Er ontstaat als het ware een ‘feedback loop’ naar de bronsystemen die het mogelijk maakt de datakwaliteit aldaar te verbeteren. “Bovendien kunnen we nu met eenduidige ‘business rules’ werken, die zorgen voor één versie van de waarheid”, meent finance manager Den

## Het gaat om een web enabled architectuur, zodat rapporten via een browser kunnen worden bekeken

Bak. “Als een consultant meent dat hij ‘lekker draait’, kunnen we dat snel uit de administratie halen. Dit betekent dat de discussie over bijvoorbeeld individuele prestaties nu op basis van feiten in plaats van meningen kan worden gevoerd. Blijkt de doorlooptijd bij een medewerker veel langer dan gemiddeld of wijkt de conversie van het aantal bezoeken in feitelijke plaatsingen substantieel af, dan kunnen we dat aan de hand van standaard rapportages of dashboards zien. Bovendien – en dat is minstens zo belangrijk – kunnen we met dit hulpmiddel nu achterhalen waar het fout gaat en dus zien aan welke knoppen we moeten draaien om tot betere prestaties te kunnen komen.”

Kamst: “Impact-analyses kunnen snel en effectief worden uitgevoerd, terwijl business-regels in de vorm van metadata eenduidig en centraal worden vastgelegd. Aanpassingen in de bronsystemen met consequenties voor het datawarehouse kunnen daarnaast op eenvoudige wijze worden doorgevoerd. Maar misschien is wel het allerbelangrijkste: het onderhoud van een analytische applicatie vraagt bij deze aanpak veel minder tijd en energie dan bij traditionele toepassingen.”

Dynalectical Solutions heeft de omgeving in vijftiendertig dagen aan Michael Page International opgeleverd. Den Bak: “Bovendien is het project zodanig aangepakt dat het voldoende schaalbaar is om naar andere landen waar Michael Page actief is uit te rollen.” Iedere dag vindt nu een incrementeel laadproces plaats. Daarbij worden vier tot vijf miljoen records geladen; dat neemt circa 2,5 uur aan tijd in beslag.

**Robbert Hoeffnagel** is freelance journalist.