

Pentaho is verrassend compleet tegen een aantrekkelijke prijs

# Veelbelovende open source BI-suite

Jos van Dongen

**Het nog steeds groeiende succes en toenemende gebruik van producten als Linux, OpenOffice en Firefox is moeilijk over het hoofd te zien. Voor tal van commerciële software zijn inmiddels verschillende open source-varianten beschikbaar. Ook in de BI-wereld wordt steeds meer gebruik gemaakt van open source-componenten.**

Een standaard Business Objects XI installatie bevat MySQL als repository database en Tomcat als applicatie-server, en SAS is bezig om de Firebird database te integreren in haar BI-platform. Dat er inmiddels ook complete open source BI-suites aan het ontstaan zijn is wellicht wat minder bekend.

De meest serieuze speler op dit terrein is Pentaho, een bedrijf dat nu twee jaar bestaat en een complete BI-suite aanbiedt. In DB/M 8 van 2005 is geschreven dat "analisten verwachten dat er op termijn een of enkele *open source gorilla's* zullen ontstaan". De vraag is of Pentaho één van deze gorilla's is; in dit artikel wordt geprobeerd met een introductie van het bedrijf en de producten alvast een eerste antwoord op deze vraag te geven.

## Geschiedenis

Pentaho bestaat sinds eind 2004 en vierde in november jongstleden haar tweede verjaardag. De mensen achter Pentaho zijn echter al veel langer actief in de BI-wereld. CEO Richard Daley en het drietal architecten James Dixon, Doug Moran en Marc Batchelor waren de drijvende krachten achter software-bedrijf Appsource dat in 1997 is overgenomen door Arbor Software (van Essbase), dat op haar beurt in 1998 weer door Hyperion is opgeslokt. Daarna is door hetzelfde team een nieuwe start gemaakt met Keyola dat vervolgens in 2002 door Lawson is opgekocht. Twee jaar geleden hebben de mannen dus hun derde start gemaakt met Pentaho, dat vooralsnog zelfstandig is. Naast dit team is André Boivert als Chairman aangetrokken (voormalig Oracle, SAS, Cognos, IBM en Sagent) en heeft Lance Walter zijn rol als VP Marketing bij Business Objects verruild voor dezelfde functie binnen Pentaho. Kortom, een door de 'BI-wol' geverfd team dat er inmiddels in is geslaagd om een solide fundament neer te leggen voor de verdere ontwikkeling van het Pentaho BI-platform. Dat er een basis is voor succes blijkt wel uit de belangstelling voor

het product, dat met 1,5 miljoen downloads in twee jaar tijd een serieuze speler geworden is in de open source-wereld.

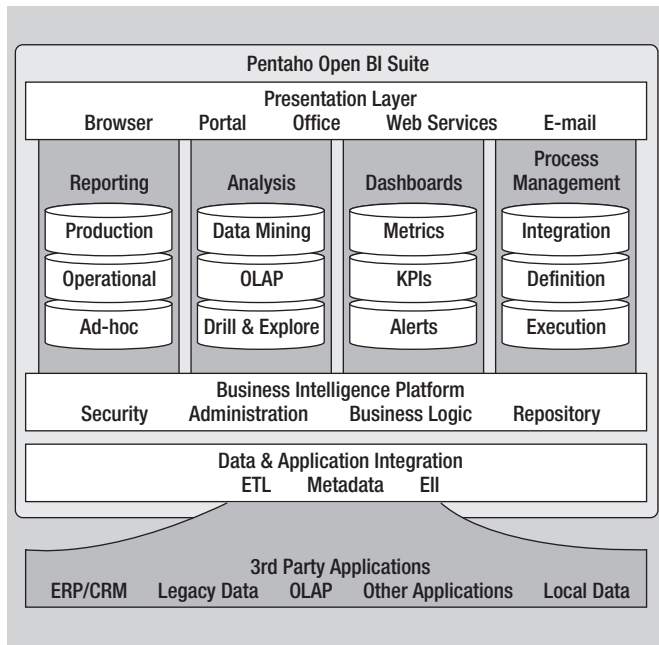
## Producten

De samenstelling van het productportfolio kan in eerste instantie een beetje diffuus overkomen. Een korte uitleg is dan ook op zijn plaats. Pentaho heeft een aantal bestaande projecten geadopteerd en deze 'suite' aangevuld met eigen producten. Het belangrijkste eigen product is het BI Platform waarin alle overige componenten zijn ondergebracht. Het platform zorgt voor interoperabiliteit, security, auditing, scheduling en distributie, en workflow-voorzieningen. De individuele producten leveren de tools voor data-integratie, reporting, OLAP, datamining en dashboards.

In afbeelding 1 is dit weergegeven. De producten zijn als complete suite beschikbaar, maar ook als losse downloads. Het voordeel van de suite is uiteraard dat dit veel ergernissen voorkomt bij installatie en configuratie. Er zijn twee varianten beschikbaar: de BI Suite (het complete platform met alle componenten) en een Reporting Server variant, dus zonder dashboards en OLAP.

De ETL-laag wordt ingevuld door het Belgische Kettle en kan ook als afzonderlijke tool worden gebruikt. Kettle zal in een vervolgartikel uitvoerig aan de tand worden gevoeld en wordt daarom hier niet verder besproken. Voordat we toekomen aan het platform zelf, wordt eerst stilgestaan bij de rapportage-, analyse- en dashboard-componenten. Deze drie onderdelen maken vrijwel altijd deel uit van de bekendere BI-tools en het is dus interessant om te bezien in hoeverre deze gratis tools vergelijkbare functionaliteit bieden. Aangezien het belangrijkste onderdeel in een BI-implementatie vaak nog de rapportages zijn zullen we deze eerst onder de loep nemen.

Als basis voor de rapportages dient het al langer bestaande JFreeReport, een op Java en XML gebaseerde rapportage-component. Oorspronkelijk is dit product ontwikkeld om



**Afbeelding 1:** Pentaho-componenten.

standaard applicaties met embedded rapportagefunctionaliteit uit te breiden, wat goed zichtbaar is wanneer we kijken naar de Advanced Reporting Guide. Deze bestaat voornamelijk uit een gedetailleerde beschrijving van de wijze waarop de XML tags gebruikt kunnen worden om een rapport op te maken, en wijkt in die zin nogal af van de standaard handleidingen van andere BI-tools.

Gelukkig heeft Pentaho inmiddels twee hulpmiddelen beschikbaar waarmee het voor normale stervelingen ook mogelijk wordt om rapporten te ontwikkelen. De eerste is een webbased report wizard waarmee in zes stappen een rapport te creëren is, de tweede is de report designer. Deze laatste is een standalone applicatie voor de ontwikkeling van jFreeReport rapporten, en doet in eerste instantie sterk denken aan MS Reporting Services, MS Access of Crystal Reports. Wanneer we echter verder kijken blijkt de functionaliteit tamelijk beperkt te zijn; er kan per rapport slechts één databron worden gedefinieerd en standaard opmaak-componenten als kruistabellen ontbreken eveneens. Weliswaar kunnen grafieken met jFreeChart worden gemaakt en vervolgens via een parameter in een rapport worden opgenomen, maar fraai is anders. Hier is dus duidelijk nog wat werk aan de winkel. Het Pentaho-platform kan echter ook met andere rapportage-tools overweg. JasperReports en BIRT (rapportage-onderdeel van het Eclipse project) worden beide ondersteund. Deze tools hebben een groter arsenaal aan standaard componenten en opmaak-mogelijkheden aan boord en zullen in de meeste gevallen dan ook de voorkeur genieten boven JFreeReport. In alle gevallen ontbreekt echter de abstractielaag die de volwassen commerciële producten kenmerkt. Voor de volgende grote release van Pentaho staan hiervoor de 'business views' op de rol, maar tot het zover is zal het op rapportagegebied nog enigszins behelpen blijven.

## Analyse

De analyse-component wordt gevormd door de Mondrian engine in combinatie met de JPivot front-end. Ook hierbij gaat het om al langer bestaande open source-projecten, waarbij Mondrian inmiddels onderdeel uitmaakt van het Pentaho-platform. Mondrian is een Java-gebaseerde OLAP server, waarmee met MDX (inderdaad, de multidimensionele query-taal die door Microsoft is ontwikkeld) analytische query's kunnen worden afgevuurd die door de Mondrian engine worden vertaald in SQL. De data zelf worden dus relationeel opgeslagen en benaderd, waarna aggregatie en calculatie door de engine wordt uitgevoerd. Feitelijk gaat het dus om een ROLAP-oplossing zoals deze bijvoorbeeld ook door Microstrategy wordt geboden. De Mondrian-architectuur bestaat uit vier lagen.

De eerste laag is de presentatielaag, waarbinnen verschillende alternatieven voor het tonen en bewerken van resultaatsets beschikbaar zijn. De tweede laag wordt gevormd door het dimensiemodel waarbinnen de MDX query wordt geparsed, gevalideerd en uitgevoerd. De derde laag is het sterschema-model waar het cachten van resultaten plaatsvindt, en de vierde laag tenslotte is de relationele opslag. Dit kan overigens elke database zijn waar een JDBC driver voor beschikbaar is. Als interactieve analyse-tool wordt zoals al aangegeven gebruik gemaakt van JPivot, een open source OLAP front-end, dat naast de Mondrian engine ook overweg kan met OLAP engines die via XMLA benaderd kunnen worden.

Het is dus mogelijk om Pentaho te gebruiken als analyse front-end voor bijvoorbeeld MS Analysis Services, Hyperion's Essbase, en SAP's Business Warehouse. Dit biedt interessante opties voor het mixen van proprietary en open source software-onderdelen. Overigens kan een JPivot analyse direct geopend worden in Excel voor eventuele verdere bewerking, en kan JFreeReport weer gebruikt worden voor het opmaken en weergeven van standaard Mondriaan OLAP-rapporten.

## Dashboards

Het onderdeel dat in eerste instantie veel potentiële gebruikers zal aanspreken is de dashboard-functionaliteit. Een dashboard kan uit diverse interactieve componenten bestaan waarbij een vorm van 'guided analysis' mogelijk wordt gemaakt. De demo-versie van Pentaho bevat het dashboard, te zien in afbeelding 2. Navigatie in het dashboard is zoals het hoort zeer intuïtief: mikken en klikken en het resultaat verschijnt in beeld. Tot zover het goede nieuws over de presentatielaag. Wanneer we onder de motorkap kijken zien we dat er geen hulpmiddelen beschikbaar zijn voor het maken van dashboards. De getoonde grafieken en lijsten dienen apart te worden gemaakt, bijvoorbeeld met JFreeReport en JFreeChart. De losse componenten moeten vervolgens in een JSP-pagina aan elkaar geknutseld worden, waarbij de interactie tussen de componenten aangebracht moet worden door de losse XML files van de juiste verwijzingen te voorzien.

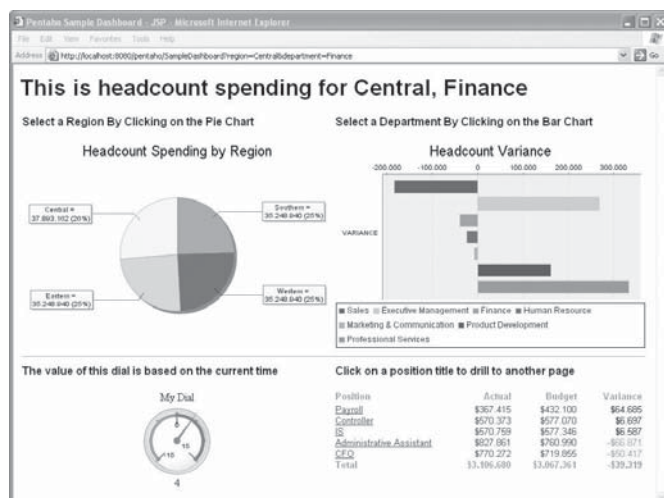
Hoewel de Designer Studio hierbij enige hulp kan bieden en er in de loop van 2007 een grafische dashboard designer gepland staat,

heeft dit weinig te maken met een standaard applicatie en lijkt het meer op gewoon maatwerk webapplicaties ontwikkelen.

## Platform

Tot slot het platform zelf, waar deze keer een prettige verrassing wacht. Met de filosofie achter het BI-platform is namelijk niets mis, en zou zelfs voor veel andere BI-leveranciers als voorbeeld kunnen dienen. Alles is namelijk actie- en procesgericht. Het idee hierachter is dat BI nooit op zichzelf kan staan maar altijd onderdeel uitmaakt van een (beslissings) proces. En als BI een onderdeel is van het proces, dan is de BI-tool dat ook. De basis van het Pentaho-platform wordt dan ook gevormd door een workflow en solution engine. Een Pentaho Solution bestaat uit een verzameling zogenaamde 'Action-Sequences' en bijbehorende documenten en definities. Voor het bouwen van een oplossing is een Designer Studio beschikbaar die beschikbaar is als plug-in voor Eclipse, de *de facto* ontwikkelomgeving voor Java-gebaseerde applicaties.

Een Action Sequence is een krachtig hulpmiddel om een veelvoud van BI-relateerde processen in te richten. Simpele acties zoals het volgens een schema of op basis van een event starten van een ETL-proces of het versturen van een e-mail kunnen worden gemaakt, maar ook complexere processen als bijvoorbeeld report bursting op basis van afwijkingen in resultaten zijn mogelijk. Een proces kan andere processen aanroepen, er is conditionele en looping logica beschikbaar en als de standaard acties niet voldoende zijn, kunnen deze met custom-acties worden uitgebreid. In het voorbeeld worden data opgehaald, en vervolgens wordt er voor elke regel in de resultaatset een mail-bericht opgemaakt en verzonden naar de bijbehorende ontvanger. Er zijn producten waar voor dit soort functionaliteit erg veel (extra) geld betaald moet worden, dus Pentaho scoort hier veel punten. Het platform bevat in de Professional Edition verder nog voorzieningen voor security, auditing, personalisatie en life cycle management. Hier wordt in een vervolgartikel meer aandacht aan besteed.



Afbeelding 2: Pentaho Dashboard.

## Uitbreiding

Vlak voor het ter perse gaan van dit nummer werd bekend dat Pentaho de rapportage-module heeft uitgebreid met webbased ad hoc query- en reporting-mogelijkheden. Dit nieuwe onderdeel maakt gebruik van de eveneens nieuwe open source metadata-laag, zodat gebruik gemaakt kan worden van zogenaamde 'Business Views' in plaats van SQL. De metadata-laag is gebouwd op basis van het Common Warehouse Metamodel (CWM). Beide onderdelen zijn gratis te downloaden vanaf de Pentaho site, zie [www.pentaho.com](http://www.pentaho.com).

## Licenties en support

Pentaho distribueert haar software in twee smaken: een open source-variant en een Professional Edition. De Professional Edition bevat dezelfde code base maar heeft daarnaast enkele opties die in bedrijfskritische omgevingen van belang zijn. Indien gekozen wordt voor de open source-variant zijn er geen licentie-kosten. Optioneel kunnen er support-contracten worden aangeschaft die gebaseerd zijn op een prijs per module (bijvoorbeeld Pentaho Reporting), per jaar en per processor. Hierbij kan gekozen worden voor Silver (\$1000,-) of Gold (\$1750) support. De Platinum support-variant maakt deel uit van de Professional Edition en bevat telefonische ondersteuning, hulp op afstand en een gegarandeerde responstijd van ten hoogste zes uur. De prijsstelling blijft met \$6.000,- per processor per jaar voor de Reporting Server nog steeds concurrerend. De professional BI Suite kent een opslag van 25 procent. Ruwweg stelt Pentaho dat de jaarlijkse licentie- en support-kosten 10 tot 20 procent bedragen van vergelijkbare 'closed source' producten. Pentaho wil echter meer zijn dan een goedkoop alternatief: de ambitie is dat klanten de producten niet kiezen vanwege de lage prijs, maar vanwege het feit dat ze beter zijn dan de concurrenten.

## Conclusie

Pentaho levert een verrassend complete BI-suite tegen een zeer aantrekkelijk prijskaartje, laat dat vooropstaan. Moeten de gevestigde partijen zich al zorgen gaan maken? Op basis van deze versie nog niet echt. Met name de rapportagefunctionaliteit is redelijk beperkt, en een omgeving inrichten en aanpassen vereist behoorlijk wat verschillende technische vaardigheden. Toch mag het product niet worden onderschat. Met de workflow- en proces-georiënteerde filosofie van het pakket is voldoende onderscheidend vermogen aanwezig om succesvol te kunnen zijn. Succes trekt vervolgens weer extra ontwikkelaars aan, waardoor een sneeuwbal-effect kan ontstaan bij het verder uitbouwen van de functionaliteit waar deze nog tekort schiet. Met de ambitieuze roadmap voor 2007 in het achterhoofd is Pentaho in elk geval volwassen en interessant genoeg om nauwlettend in de gaten te houden.

Jos van Dongen ([jvdongen@tholis.com](mailto:jvdongen@tholis.com)) is Senior Consultant bij Tholis Consulting.